

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 00:30:12 : Search time 66.4937 Seconds
(without alignments)
10792.671 Million cell updates/sec

Title: US-09-698-295-10
Perfect score: 14333
Sequence: 1 MWSEEEEDGDAEETQDSE.....KLGPKASRSHNNKLOSTAS 2781

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriaph:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14333	100.0	2781	4 Q9UIG2	Q9UIG2 homo sapien
2	3535	24.7	796	11 Q8CFX5	Q8CFX5 mus musculus
3	3445	24.0	803	11 Q8K092	Q8K092 mus musculus
4	2848.5	19.9	2649	5 Q9W0T2	Q9W0T2 drosophila
5	2842.5	19.8	2669	5 Q9W0T1	Q9W0T1 drosophila
6	2839.5	19.8	2669	5 Q95VB8	Q95VB8 drosophila
7	2775.5	19.4	645	11 Q8VDN7	Q8VDN7 mus musculus
8	1368.5	9.5	412	4 Q9H5E0	Q9H5E0 homo sapien
9	871.5	6.1	1711	5 Q45409	Q45409 caenorhabdi
10	865.5	6.0	1713	5 Q95Z08	Q95Z08 caenorhabdi
11	674.5	4.7	1022	5 Q960Y3	Q960Y3 drosophila
12	580.5	4.1	7210	5 Q9V7G8	Q9V7G8 drosophila
13	580.5	4.1	9270	5 Q8MLD9	Q8MLD9 drosophila
14	572.5	4.0	2768	5 Q9VC00	Q9VC00 drosophila
15	572	4.0	17352	5 Q95YK2	Q95YK2 procamburus
16	560	3.9	5476	5 Q9NJ17	Q9NJ17 drosophila

ALIGNMENTS

RESULT 1

Q9UIG2
ID Q9UIG2 PRELIMINARY; PRT; 2781 AA.
AC Q9UIG2
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Bromodomain PHD finger transcription factor.
GN BPTF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=20130111; PubMed=10662542;
RX Jones M.H., Hamana N., Shimane M.;
RT "Identification and characterization BPTF, a novel bromodomain
transcription factor.";
RL Genomics 63:35-39(2000).
DR EMBL; AB032251; BAA89208.1;
DR HSP; Q92831; 1B91.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT_dom.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR001963; Znf_PHD.
DR Pfam; PF004439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 2.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 2781 AA; 311210 MW; 7B3BE0C1C4C3CFBE CRC64;

Query Match

100.0%; Score 14333; DB 4; Length 2781;

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2781; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MYSEEEEDGDAEETQSEDEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHST	60
Db	1	MYSEEEEDGDAEETQSEDEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHST	60
Qy	61	YSSTPGRRKPRVHRPRPILPEKIDPPEPPKSSDLMVNEHIMNVIAIYEVLRNFGTV	120
Db	61	YSSTPGRRKPRVHRPRPILPEKIDPPEPPKSSDLMVNEHIMNVIAIYEVLRNFGTV	120
Qy	121	LRLSPFFEDFCAALVSOEQCTLAEMHVLLKAVLEEDTSNTTGGPADLKDSVNSTLY	180
Db	121	LRLSPFFEDFCAALVSOEQCTLAEMHVLLKAVLEEDTSNTTGGPADLKDSVNSTLY	180
Qy	181	FIDGWTWPEVLRVYCESDKYHHVLPYQEAEDYPYGVENKIKVLQFLVQFLTNTIARE	240
Db	181	FIDGWTWPEVLRVYCESDKYHHVLPYQEAEDYPYGVENKIKVLQFLVQFLTNTIARE	240
Qy	241	ELMSEGVIOYDDHCRVCHKLDLCCETCSAVYHLECVKPLEBVPDEMOCEVCVAHKV	300
Db	241	ELMSEGVIOYDDHCRVCHKLDLCCETCSAVYHLECVKPLEBVPDEMOCEVCVAHKV	300
Qy	301	PGVTDCAEIQKKNPYIRHEPIGYDRSRKRYWFLNRRLIIEEDTENENEKKIWIYSTKVQ	360
Db	301	PGVTDCAEIQKKNPYIRHEPIGYDRSRKRYWFLNRRLIIEEDTENENEKKIWIYSTKVQ	360
Qy	361	LAELIDCLDKDYWAELCKLILEENREITHRHMDITEDLTNKARGSNKSFLLAANEELLES	420
Db	361	LAELIDCLDKDYWAELCKLILEENREITHRHMDITEDLTNKARGSNKSFLLAANEELLES	420
Qy	421	IRAKKGIDINVKSPEETKDNKTENDSKDAEKREEFEDQSLKSDDKTDPDDPEQK	480
Db	421	IRAKKGIDINVKSPEETKDNKTENDSKDAEKREEFEDQSLKSDDKTDPDDPEQK	480
Qy	481	SEVGDFKSEKSNGLSESPGAGKASGSTRITILRNPDLSKLQSKQVAAAHAENKL	540
Db	481	SEVGDFKSEKSNGLSESPGAGKASGSTRITILRNPDLSKLQSKQVAAAHAENKL	540
Qy	541	FKEKEVLVNSQGEISRLSTKKEVIMKGNINNYFKLQEGKYRVYHNOYSTNFALNKH	600
Db	541	FKEKEVLVNSQGEISRLSTKKEVIMKGNINNYFKLQEGKYRVYHNOYSTNFALNKH	600
Qy	601	QHRDHDKRHLAHKFLTPAGEFKWGSVHGSKVLITSLRLITITOLENNIPSFLLHPN	660
Db	601	QHRDHDKRHLAHKFLTPAGEFKWGSVHGSKVLITSLRLITITOLENNIPSFLLHPN	660
Qy	661	WASHRANWIKAVQMSKPREFALALILECAVKPVWMLPIWREPLGHTRLHMTSIEREE	720
Db	661	WASHRANWIKAVQMSKPREFALALILECAVKPVWMLPIWREPLGHTRLHMTSIEREE	720
Qy	721	KEKYKKEKKOEETQQAATWVKYTPPVKHQVWKQGEERYVTGYGWSWISKTHVYRF	780
Db	721	KEKYKKEKKOEETQQAATWVKYTPPVKHQVWKQGEERYVTGYGWSWISKTHVYRF	780
Qy	781	VPKLPNGTNVYRKSLSTGNMNDENDESCKRSRPPKIKIETPSEKDEVKGSAAK	840
Db	781	VPKLPNGTNVYRKSLSTGNMNDENDESCKRSRPPKIKIETPSEKDEVKGSAAK	840
Qy	841	GADQNMDSKITEKKQDQVKELLDSDSKPEEPMEVDMDKTESHVNCQESSQVDVY	900
Db	841	GADQNMDSKITEKKQDQVKELLDSDSKPEEPMEVDMDKTESHVNCQESSQVDVY	900
Qy	901	NVSEGFHLRTSYKKKTSKSLDGLLERRIKOFTLEEKORLEKIKLEGKIGKIGTSTNSS	960
Db	901	NVSEGFHLRTSYKKKTSKSLDGLLERRIKOFTLEEKORLEKIKLEGKIGKIGTSTNSS	960
Qy	961	KNLSESPVITAKEGGQSDSMROQSPNANNQDPEDLIQGCSQSDSSVLRMSDPSHTNK	1020
Db	961	KNLSESPVITAKEGGQSDSMROQSPNANNQDPEDLIQGCSQSDSSVLRMSDPSHTNK	1020
Qy	1021	LYPKDRVLDVYSIRSPETKCPKQNSIENDIEEKVSDLASRGOEPTKSKTKGNDFFIDDSK	1080
Db	1021	LYPKDRVLDVYSIRSPETKCPKQNSIENDIEEKVSDLASRGOEPTKSKTKGNDFFIDDSK	1080
Qy	1081	LASADDICTLCKNKKPLIQEESDTIVSSSKSALHSSYPKSTNDRDATPLSRAMDFEGL	1140
Db	1081	LASADDICTLCKNKKPLIQEESDTIVSSSKSALHSSYPKSTNDRDATPLSRAMDFEGL	1140
Qy	1141	GCDSESNLSNSDTSVSIQDSSSEDMIVQNSNESISEQFRTREODVLEPLKCELVS	1200
Db	1141	GCDSESNLSNSDTSVSIQDSSSEDMIVQNSNESISEQFRTREODVLEPLKCELVS	1200
Qy	1201	ESTGNCEDRLPVKGTGANGKPKSQOKLEERPVNKCSDQIKLNTTDRKNNENRESEKKG	1260
Db	1201	ESTGNCEDRLPVKGTGANGKPKSQOKLEERPVNKCSDQIKLNTTDRKNNENRESEKKG	1260
Qy	1261	QRTSTFQINGDKNPKIYLKGECLKEISESRVSVNVEPKVNNINKIIPENDIKSLTVKE	1320
Db	1261	QRTSTFQINGDKNPKIYLKGECLKEISESRVSVNVEPKVNNINKIIPENDIKSLTVKE	1320
Qy	1321	SAIRPFINGDVIMEDFERNESSETKSHLLSSDABGNYSRDSLETLPSPKESDSTOTTPS	1380
Db	1321	SAIRPFINGDVIMEDFERNESSETKSHLLSSDABGNYSRDSLETLPSPKESDSTOTTPS	1380
Qy	1381	ASCPESNVQVQVEDMEIETSEVKKVTSSPITSEESNLNDFIDENGLPINKNNENNGES	1440
Db	1381	ASCPESNVQVQVEDMEIETSEVKKVTSSPITSEESNLNDFIDENGLPINKNNENNGES	1440
Qy	1441	KRKVTITVETVMTSTVATESKTVIKVEKGDQFVVSSTENCAKSTVTTTTVTIKLSTPS	1500
Db	1441	KRKVTITVETVMTSTVATESKTVIKVEKGDQFVVSSTENCAKSTVTTTTVTIKLSTPS	1500
Qy	1501	TGGSVDIISVKEQSKTVVTVTDSLTGTGTLVTSMTVSKYSTRDKVKLMPKSRPKKT	1560
Db	1501	TGGSVDIISVKEQSKTVVTVTDSLTGTGTLVTSMTVSKYSTRDKVKLMPKSRPKKT	1560
Qy	1561	RSGTALPSYRKPFVTKTKSIFVLNDDLLKLRKGGIREVPYFNNAKPAIDIMPYPSP	1620
Db	1561	RSGTALPSYRKPFVTKTKSIFVLNDDLLKLRKGGIREVPYFNNAKPAIDIMPYPSP	1620
Qy	1621	RPTFGITWRYLQVTKVSLAGVSLMURLLWASLRWDDMAKVPVGGSGTSTETSEITITT	1680
Db	1621	RPTFGITWRYLQVTKVSLAGVSLMURLLWASLRWDDMAKVPVGGSGTSTETSEITITT	1680
Qy	1681	EIIKRRDVGPIREYCIIRKICIGVPEPKETPTORGLRSALRPKRPETPKOTG	1740
Db	1681	EIIKRRDVGPIREYCIIRKICIGVPEPKETPTORGLRSALRPKRPETPKOTG	1740
Qy	1741	PVIETWABEELELWEIRAFERKEKAQAVEQAKRLEQKQPTVIATSTTSST	1800
Db	1741	PVIETWABEELELWEIRAFERKEKAQAVEQAKRLEQKQPTVIATSTTSST	1800
Qy	1801	TSTISPAOKVAVPISGVTGTGMVLTQVGSPTATVFOONKNPHOTFATVWKGQNS	1860
Db	1801	TSTISPAOKVAVPISGVTGTGMVLTQVGSPTATVFOONKNPHOTFATVWKGQNS	1860
Qy	1861	GWQVQVQVGLIISSTGTSOOTFTSPORATVIRRENTSGSGGTTNSOVITGPQIRP	1920
Db	1861	GWQVQVQVGLIISSTGTSOOTFTSPORATVIRRENTSGSGGTTNSOVITGPQIRP	1920
Qy	1921	GMTVIRTPLOQSTLGKAIIRTPVMVQPCAPQVMTQIIRGQPVSTAVSAPNTVSTPGOK	1980
Db	1921	GMTVIRTPLOQSTLGKAIIRTPVMVQPCAPQVMTQIIRGQPVSTAVSAPNTVSTPGOK	1980
Qy	1981	SLTSATSTNSIQSSASQPPRQOQVQKLTMAQLTQLTQGHGNGQLTVVIOGQGTQOL	2040
Db	1981	SLTSATSTNSIQSSASQPPRQOQVQKLTMAQLTQLTQGHGNGQLTVVIOGQGTQOL	2040
Qy	2041	QLIPQVTVLPGCQQLMQAAMPNGTVORFETPLATTATTAATTTTSTVTAAGTGBOR	2100
Db	2041	QLIPQVTVLPGCQQLMQAAMPNGTVORFETPLATTATTAATTTTSTVTAAGTGBOR	2100
Qy	2101	QSKLSPQMVHODKTLTPPAQSSSVGPAKAQPTAOPSARPOPTOPQSPAQPEVOTQPEV	2160
Db	2101	QSKLSPQMVHODKTLTPPAQSSSVGPAKAQPTAOPSARPOPTOPQSPAQPEVOTQPEV	2160

QY 2161 QTQTVSHVSEAPTHAQSSKPVAAQSQPNVQSQSPVRVQSPQTRIRPSTPSQL 2220
DB 2161 QTQTVSHVSEAPTHAQSSKPVAAQSQPNVQSQSPVRVQSPQTRIRPSTPSQL 2220
QY 2221 SPGQSQVQTTSTQPIQPHSTSIQIPSGQSQPOQVQSQSTOTLSSGOTLNQVSVSSPS 2280
DB 2221 SPGQSQVQTTSTQPIQPHSTSIQIPSGQSQPOQVQSQSTOTLSSGOTLNQVSVSSPS 2280
QY 2281 RPQIQIQPOQVITAVPOLQOQVQLSQIQSQVVAQIQAAQSGVPPQIKLQPLQIQOQS 2340
DB 2281 RPQIQIQPOQVITAVPOLQOQVQLSQIQSQVVAQIQAAQSGVPPQIKLQPLQIQOQS 2340
QY 2341 AVQTHQIQNVVYVQAASVQEQQLRVQQLRDQOQKKQOQIIEIKREHTLQASNOSEIQK 2400
DB 2341 AVQTHQIQNVVYVQAASVQEQQLRVQQLRDQOQKKQOQIIEIKREHTLQASNOSEIQK 2400
QY 2401 VVMKHNAVIEHLKOKKSMTPAEREENQRMIVCNQVMKYILDKIDKEEKQAAKKRREESV 2460
DB 2401 VVMKHNAVIEHLKOKKSMTPAEREENQRMIVCNQVMKYILDKIDKEEKQAAKKRREESV 2460
QY 2461 EOKRSKONATKLSALLFKHKLRAELTKRALLDKDLQIEVQBELKRDLLKKEKDLQ 2520
DB 2461 EOKRSKONATKLSALLFKHKLRAELTKRALLDKDLQIEVQBELKRDLLKKEKDLQ 2520
QY 2521 LAQATAVAACPVPVLPAPAPPPPPPPPPVQHTGLJSTPTLPVASQKRKREEEKDS 2580
DB 2521 LAQATAVAACPVPVLPAPAPPPPPPPPPVQHTGLJSTPTLPVASQKRKREEEKDS 2580
QY 2581 SSKSKKKMSTTSKTKKOTKLYCICKTPYDESKEYIGCDRCQNWYHGRVCGVILQSEAE 2640
DB 2581 SSKSKKKMSTTSKTKKOTKLYCICKTPYDESKEYIGCDRCQNWYHGRVCGVILQSEAE 2640
QY 2641 LIDYVCPQOQSTEDAMTVLTPLEKDYEGKRLVLSLQAHKMAWPLEPVPDNDADPY 2700
DB 2641 LIDYVCPQOQSTEDAMTVLTPLEKDYEGKRLVLSLQAHKMAWPLEPVPDNDADPY 2700
QY 2701 GVKEPMDLATMERVORRYEKELEFVADNTKIFDNCRYNPNDSFPYQCAEVLESFFV 2760
DB 2701 GVKEPMDLATMERVORRYEKELEFVADNTKIFDNCRYNPNDSFPYQCAEVLESFFV 2760
QY 2761 QKLGFKASRSHNNKLOSTAS 2781
DB 2761 QKLGFKASRSHNNKLOSTAS 2781
RESULT 2
Q8CFX5 PRELIMINARY; PRT; 796 AA.
AC Q8CFX5
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE Similar to fetal alzheimer antigen (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RA Strausberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC037661; AAH37661.1; -
FT NON_TER 796 796
SQ SEQUENCE 796 AA; 90592 MW; 2C1BBE28009C82F3 CRC64;
Query Match 24.7%; Score 3535; DB 11; Length 796;
Best Local Similarity 84.3%; Pred. No. 9e-144;
Matches 671; Conservative 35; Mismatches 24; Indels 66; Gaps 4;
QY 1 MVSEDEEEEDGAEETQDSEDEEDDEEDDDSDYPEEMEDDDDDASYCTESSFRSH 58
DB 1 MVSEDEEEEDGAEETQDSEDEEDDEEDDDSDYPEEMEDDDDDASYCTESSFRSH 58

DB 1 MVSEDEEEEDGAEETQDSEDEEDDEEDDDSDYPEEMEDDDDDASYCTESSFRSH 60
QY 59 STYSTPGRKRPRVHRPRSPILLEKDIPLPEPKSSEDLMPNEHIMNVIAIYEVLRNFG 118
DB 61 STYSTPGRKRPRVHRPRSPILLEKDIPLPEPKSSEDLMPNEHIMNVIAIYEVLRNFG 120
QY 119 TVLRUSPFRFEDFCALYSQEQCTLMAEMHVVLKAVLREEDTSNTTGPADLKDSVNST 178
DB 121 NVLRUSPFCFEDFCALYSQEQCTLMAEMHVALLKAVLREEDTSNTTGPADLKDSVNST 180
QY 179 LYFIDGWTWPEVLRVYCESDKKEYHHVLPYQAEADYPGYGVENKIKVLPFLVDQFLTTNIA 238
DB 181 LYFIDGWTWPEVLRVYCESDKKEYHHVLPYQAEADYPGYGVENKIKVLPFLVDQFLTTNIA 240
QY 239 REELMSEGVIOYDDHCRVCHKIGDLCCETCSAVYHLECVKPPLEEVPEDEQCEVCVAH 298
DB 241 REELMSEGVIOYDDHCRVCHKIGDLCCETCSAVYHLECVKPPLEEVPEDEQCEVCVAH 300
QY 299 KVPGVTDCAVEIQKNKPIIRHEPIGYDRSRKRYWFLNRLIITEEDTENENKKIYYSTK 358
DB 301 KVPGVTDCAVEIQKNKPIIRHEPIGYDRSRKRYWFLNRLIITEEDTENENKKIYYSTK 360
QY 359 VOLAELIDCLDKDYWEAELCKTILEMREEIHRHMDITEDLTNKGSKNSFLAAANEEL 418
DB 361 VOLAELIDCLDKDYWEAELCKTILEMREEIHRHMDITEDLTNKGSKNSFLAAANEEL 420
QY 419 ESIRAKKG-DIDNVKSPSETEKDKNETEND-SKDAENREEFEDQSLKSDDKTPDDDP 476
DB 421 DSLRIKRGEDIDCQSPEDPEKDHGEGENSSKDAEKSEAEADPSADKDADSKGLEEP 480
QY 477 EQGKSE----- 482
DB 481 GHGKPEEPTGYDGNKSNVPA NLGNTTNASPEETSPCDGRSPGCLSETHDSSSMAEKKV 540
QY 483 -----VGDFKSKNSGELSESAGKAGSGSTRITIRLNPDSKLSQKSOVAAAA 534
DB 541 ASELPPDPVPGDFRLEKSNGEVSESPGCKGTSGSTRIITIRLNPDSKLSQKSOVAAAA 600
QY 535 HEANKLPKEGKVLVNVNQSGEISRLSTKKEVIMKNNINNYFKLGQEGKYRVYHNOYSTNS 594
DB 601 HEANKLPKEGKVLVNVNQSGEISRLSTKKEVIMKNNINNYFKLGQEGKYRVYHNOYSTNS 660
QY 595 FALANKHOHREDHDKRHLAKFCLTPAGEFKWNGSVHGSVLTISTLRITITOLENNIPS 654
DB 661 FALANKHOHREDHDKRHLAKFCLTPAGEFKWNGSVHGSVLTISTLRITITOLENNIPS 720
QY 655 SFLEPNWASHRANWIKAVQMSKPREFALALALECAVKPVVMLPIWREFLGHTRHMT 714
DB 721 SFLEPNWASHRANWIKAVQMSKPREFALALALECAVKPVVMLPIWREFLGHTRHMT 780
QY 715 SIEREEKKVKKKKK 730
DB 781 SIEREEKKVKKKKK 796
RESULT 3
Q8K092 PRELIMINARY; PRT; 803 AA.
AC Q8K092;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC032289; AAH32289.1; -

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 01:17:18 ; Search time 109.881 Seconds
(without alignments)
3939.095 Million cell updates/sec

Title: US-09-698-295-1

Perfect score: 14971

Sequence: 1 MVSEEEEDGAEETQDSE.....KLKGFKASRSHNNKLOSTAS 2907

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 556269 seqs, 148893369 residues

Total number of hits satisfying chosen parameters: 556269

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	9.3	328	9	US-09-925-297-816
2	982.5	6.6	238	9	US-09-729-835-80
3	456.5	3.0	3664	15	US-10-177-293-423
4	443	3.0	2344	9	US-09-815-242-12713
5	438	2.9	3256	10	US-09-919-172-98
6	432	2.9	3256	11	US-09-919-039-21
7	432	2.9	2665	9	US-09-864-761-34248
8	402	2.7	4019	10	US-09-738-973-425
9	402	2.7	4019	10	US-09-854-133-425
10	402	2.7	4019	15	US-10-144-649A-425
11	401.5	2.7	1367	10	US-09-801-368-108
12	391.5	2.6	3899	15	US-10-171-311-4
13	391.5	2.6	3917	15	US-10-171-311-8
14	391	2.6	5179	9	US-09-922-217-1068
15	391	2.6	5179	10	US-09-833-263-1068

16	391	2.6	5179	14	US-10-025-380-1068	Sequence 1068, Ap
17	389.5	2.6	2478	9	US-09-815-242-5816	Sequence 5816, Ap
18	389.5	2.6	2478	9	US-09-815-242-12967	Sequence 12967, A
19	387.5	2.6	3907	15	US-10-171-311-2	Sequence 2, Appli
20	387.5	2.6	3925	15	US-10-171-311-6	Sequence 6, Appli
21	384	2.6	5935	15	US-10-243-243A-8	Sequence 8, Appli
22	381.5	2.5	5877	15	US-10-142-515-11	Sequence 11, Appl
23	379	2.5	2701	15	US-10-171-311-83	Sequence 83, Appl
24	373	2.5	2476	11	US-09-824-574-7	Sequence 7, Appli
25	370	2.5	2297	15	US-10-245-802-20	Sequence 20, Appl
26	362.5	2.4	2843	10	US-09-987-482-1	Sequence 1, Appli
27	361.5	2.4	2843	8	US-08-681-219-32	Sequence 32, Appl
28	361.5	2.4	2843	12	US-10-092-138-30	Sequence 30, Appl
29	354.5	2.4	1400	10	US-09-764-176-7	Sequence 7, Appli
30	344.5	2.3	2382	15	US-10-196-935A-2	Sequence 2, Appli
31	342	2.3	2368	9	US-09-815-242-5635	Sequence 5635, Ap
32	342	2.3	2368	9	US-09-815-242-12389	Sequence 12389, A
33	339	2.3	6281	9	US-09-815-242-12996	Sequence 12996, A
34	337	2.3	528	12	US-09-840-746-20	Sequence 20, Appl
35	331	2.2	1236	11	US-09-769-787-109	Sequence 109, App
36	329.5	2.2	5795	9	US-09-815-242-12610	Sequence 12610, A
37	328	2.2	1863	10	US-09-734-672-4	Sequence 4, Appli
38	328	2.2	1863	11	US-09-982-828-6	Sequence 6, Appli
39	324	2.2	1781	9	US-09-738-877-3	Sequence 3, Appli
40	324	2.2	1781	11	US-09-961-403-13	Sequence 13, Appl
41	322	2.2	1863	10	US-09-734-672-2	Sequence 2, Appli
42	322	2.2	1863	10	US-09-734-672-6	Sequence 6, Appli
43	322	2.2	1863	11	US-09-982-828-2	Sequence 2, Appli
44	322	2.2	1863	11	US-09-982-828-4	Sequence 4, Appli
45	322	2.2	1863	15	US-10-022-819-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-925-297-816
; Sequence 816, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 816
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (170)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (172)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (174)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (178)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (183)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (269)

Db 787 KNEKTDKERTFDPVERERRLRKEKVEKDKTKQ----- 822
Qy 60 TYSPTGRRKPRVHRSPILE-----EKDIPPLEFPKS-----SDLMVPNEHI 104
Db 823 -----KRGKVHSPSSQSETQENEREOSP-EKPRSCNKLSEKADKEGIAKNRLE 874
Qy 105 MNVIAIYVLRNFTVRLSPFREDCAALVSQEOCTLMAEMHVLLKAVLREEDTSNT 164
Db 875 MPCVVULTRVKEGKVIDHTPV-----EKLKALDNDTVKSSALDQKLQVSOT 922
Qy 165 TFGPADLK--DSVNSTLYFIDGM-TWPEVLRVYCESD---KEYHHVLPYQEAEDPYGPV 218
Db 923 EPAKSDLSKLESVRMKPKGKLSHVVEV---EKEGRUKARKHLAPEQAD---GVS 974
Qy 219 ENKIKVLOFLVDQFLTNIAAREELMSE---GVIOYDDHCRVCHKLGLDLCCTCSAVYHL 275
Db 975 AVDLKLEAKRRRREADNLAKAEKQKPEVKKSSPEMED-ARVLSKKQPDVSSREVILLREG 1033
Qy 276 ECVKPPIL--EVPDEQOCEVCVAHKVPYGVTDCAEIOKNKPYIRHPEPIGYDRSRKYWF 333
Db 1034 EAERKPKRKETLKRESKKILDRLNTVAPSKDC-OELAS-----ISVSGSRPSSD 1083
Qy 334 LNRRIL--IIBEDTENENEKKIWIYYSYTKVQLAELIDCLDKDYEAELCKILEEM---REEI 388
Db 1084 LQARLGLAGESVENEQVSKKPIPSKPLKQL-----QVLDQGPREDV 1129
Qy 389 HR-HMDJTEJLTNKGARNSKFLAAANEIILESIRAKKG-DIDNVKS---PEETEKDNE 443
Db 1130 RKNYCSLRDPTPERKSGQESHVSNTTEKI-----GIDIDHTQSYRQMEQSRKQ 1181
Qy 444 TENDSKDAE-----KNREFEQDGL---EKSDDKTTPDDDPQOGESEPTVEG--- 488
Db 1182 MEMBIAKSEFGSKPKDQDYERKSLVHEVCKPQDVTDDSPSKKKRMDHVDICTKR 1241
Qy 489 DKGNSVANLGDNTNATSEET--SPS-----EGRSPVGC--LSETPDSSNMAEKVA 537
Db 1242 ERNYSRQISED-----SERTGSPSVRHGSPHEDEDPICSPLLSVKSGPKVDEK--- 1293
Qy 538 SELPOD---VPEEPNK--TCESSNTSATTSIOPLNLSNSSLSSQSESASAKA--- 588
Db 1294 -VLPYSNITVREESLKENFYDSSRREQMAQMAKIKLSVLNSEDLELNRWDSQMKDAGRFD 1352
Qy 589 -----DOPENCERESHPTVSIQOEIVGDFTSEKSTGE 620
Db 1353 VSPNSIIRKDSLRKVRDLEGEVPSDDEGEHKSHP-----RASA 1397
Qy 621 LSESPGAGKAGSGSTRIITRLRNPDSLSOLKSQOAAAAHAANKLFKEG-KEYLVVNSQ 679
Db 1398 LYE-----SSRLSFLLRDREDKLRE-RDERLSSSL-ERNKFYSFALDKTITPDTK 1445
Qy 680 GEISR--LSTKKEVIMKGINNINNYFKLGQEGKYRVYHNYQ----- 716
Db 1446 ALLERAKLSSSRE-----ENWSFLDWSRPFANFRNNKDKKVDASAPRPIPSWYMKK 1498
Qy 717 -----SINSFALNKHQREHDHRRHL-AHKFCLTTPAGEFKWNGSVGKVLITSTLRLT 770
Db 1499 KIRTDSEGMDDKEDKHEQEERQELFASRF-----LH-SSIFEODSKRL- 1543
Qy 771 ITOLENNIPSSFF-----HPNASHRANWIKAVQMCSPREFALALAILCAVKKPVVMLP 825
Db 1544 -QHLEKREEDSDIFSIRIYGKQTSSEGAN-----STTDSIQEPPVILF- 1583
Qy 826 IWREFLGTHLRMTSITEREEKVKKKKEKEEETMOQATWYKTFPVKQHWQKGE 885
Db 1584 -----HSRFMELTRMQOKEKEKQKPEVEQEDTENHP-----KTPESAPENKDS 1629
Qy 886 EYRV---TGVGGSWSISKTHYRVFKLPNGTNTVNYKKSLEG---TKNNMDENMDESKR 939
Db 1630 ELKTPPSVGPSPVTVVLTLESAPSALERTTGD-----KTVEAPLVTEKTEVEPATVSEE 1683
Qy 940 KCSRSPKIKIEPDSEKDEKVGSDAAKA-----DQ----- 970
Db 1684 KPASEPAPVQLEQVDLPFGADPKDEKAAMMPAGVEEGSGQPPYLDKAPPTPGASF 1743

Qy 971 -----NEMDISKITEKKQDVVKELLDSDSKFCKEPEMEVDDDMKTESHVNCQ-----E 1019
Db 1744 QAESNVDPDSTQPLSKPAQKSEANEPAEKDPATADAPDANQAKAAEAPESOPPASE 1803
Qy 1020 SSQVDVNVSGFHLRTSYKKKT-----KSSKLDGLLERIKQFTLE-- 1061
Db 1804 DLEVDPVPAAKDKKPNKSKRSKRTVQAAAVSIVEKPVYTRKSERID---REKLKRSNSPRG 1860
Qy 1062 EKORLEKIKLGGGKIGKGTSTNSKNLSESPVITKAKEGCOSQSMROEQSPNANDPOE 1121
Db 1861 EAQKLLLEKWEA--EKTRTASKNSAADLEHPEPSLPLSRTRRRNVRVSYYATMGDHENRS 1918
Qy 1122 DLIOGCSQSQSSSVLMSDPSHTTNKLPKRVLDVDSI-----RSPETKCPKQNSIE 1173
Db 1919 PVKEPVEQ-----PRVTRKRL---ERELQEAAPVTPTRGRPKPTRRADEBEE 1965
Qy 1174 NDIEEKVSDL--ASRGOEPKSKT-----KG-NDFFIDDSKLASADDIGLICNKK 1221
Db 1966 NEAKEPAETLKPPEGWRSRPSQKTAAGGGPOCKKGKNEPKVDATREPAATTEVGPIGVKE 2025
Qy 1222 KPLIOEESDITIVSSSKSALHSSVPKSTNDRDATPLSRAMDPEGLGCDSESNSTLENSSD 1281
Db 2026 SSMEPKAAEEBAGSEQKDRKDRKAGTDKNPPTAPVEVV--EKKPAPEKNKSKRGRSRN 2082
Qy 1282 TVSIQDSSEEDMIYONSNESTIS-----EQFRTREQDVEVLEPLKCELVSGESTGNCEDRL 1336
Db 2083 SRLAVDKSAS---LKNVDAAVPRGAAQAAGERESGVVAVSPEKSE----- 2125
Qy 1337 PVKTEANGKKPSOOKLEERPKNKCSQIKLNTYDKNNENRESEKKGORTSTFQING 1396
Db 2126 -----SPOKE-----DGLSQLKSDVPDPKPEKEDVDVSAGSPSEATQL-- 2165
Qy 1397 KDNKPKIYLGECLKEISESRVSGNVPEKPVNNINKIIPENDIK-----SLTVKESATRP 1451
Db 2166 ---AKOMLEQAEVHEHLAKLAEASASAAYKADAPGLAPEDRDRPAHQASETELAAGAIGS 2221
Qy 1452 FINDGVIMEDFN-----ERNSSETKSHLLSSSDAEG 1482
Db 2222 IIN-DISGEPENFAPPYPYPGESQTDLQPPAGAQAALQPSERGMEDEAVSGILETEAATE 2280
Qy 1483 NYRDSLETL-PS-----TKESDSTQTTTPSASCPSNSVNOVEDMEITSEVKKVTSSPI 1536
Db 2281 SSRPPVNAPODSAGTDTKEAGNSSET-SHSVPEAKSGKEVEVTLVRKDKRGOKTT--- 2336
Qy 1537 TSEESNLNDFIDENGLPINKNENVNGESKRKTVITEVTTMTSTVATESKTIVKVEKGD 1596
Db 2337 RSRKRNTNKKVAPVESHVPESNOAOGES---PAANEGETTVQHPPEAQE-----EKQS 2387
Qy 1597 KQTVVVSSTENCASQSTVTTTTVTKLSTPSTGGSDIISVKEQSKTVVTTTTVTDLSLTG 1656
Db 2388 EKPHSTPQSC-----TSDLSKI--PSTENSQEISVEERTPT--KASVPPDLPPP- 2434
Qy 1657 GTLVTSMVTSKEYSTRQVKMLKFSRPKKTRSGTALPSYKRVTKSKTKSFVLPNDLLK 1716
Db 2435 ---PQAPVDEEPOARFVHSIIESDPVTPPSDSP------IPTLPSVTA 2478
Qy 1717 KLA---RKGKIREVPYENYNAKPAIDIWVPSPRPTFGITWRYLQTVKSLAGVSLMLRL 1773
Db 2479 KLSPPVASGGI-----PHQSP--PTKVTEWITROEPEPRAOSTFS----- 2515
Qy 1774 LWASLRWDDMAAKVPPGGGSTR-----TETSETTEITTEIKRRDVGPGYIRFEYCIRK 1827
Db 2516 --PALPDDTKASDVTSSSTLRLKILMDPKVYSATSVTSTSV-----TTA 2557
Qy 1828 IICPIGVPEPKETPTPQKRLRSALRPKRPETPKQTPGVIIETWVAEELWEIRAF 1887
Db 2558 IAEVSAAPCLHEAPPP-----PVDSKKPLEKTEAPPV-----TNNSEIOASELVA 2604
Qy 1888 AERVEKEKAQAV-----EQQAQKRLKQOKPTVIATSTPSTSTSTST 1929
Db 2605 A---DKEKAPVAPKITSVSRMPVSIDLENSOKITLAKPAPOTLGLVSLGTLVNV 2661

RESULT 4	
US -09-815-12713	
; Sequence 12713, Application US/09815242	
; Patent No. US20020061569A1	
; GENERAL INFORMATION:	
; APPLICANT: Haselbeck, Robert	
; APPLICANT: Ohlsen, Kari L.	
; APPLICANT: Zyskind, Judith W.	
; APPLICANT: Wall, Daniel	
; APPLICANT: Trawick, John D.	
; APPLICANT: Carr, Grant J.	
; APPLICANT: Yamamoto, Robert T.	
; APPLICANT: Xu, H. Howard	
; TITLE OF INVENTION: Identification of Essential Genes in	
; PROKARYOTES	
; FILE REFERENCE: ELITRA.011A	

Db 266 LQDYATEKESADGLQ-----GETQLL-----VSRKSRPKSGSGHAVA 304
QY 411 AAAN--EETILESIRAKKGIDINWKPEET-----EKDKNET-----ENDSKDAEKNR 455
Db 305 BPASPEQELDONKGRDVESQVTPSKAVGASFLPIYEPAKMKTTPVQVISOQONSQKHKNK 364
QY 456 EEF-----EDOSLEKD-----SDDKTPDDDPQKSGSEETPYGDKGNSVY--ANLGDNTN 504
Db 365 DLYTTGRRESYNLGKSEFGKAGDKTLTPRKLSTRNRTPAKVEDAADSAATKPNLSSKTRG 424
QY 505 A--TSEETSPB--GRSPVCLSET-----PDSSNMAEK-----KVASLPODYPE 546
Db 425 SIPTDVEYLPTEIETIHNPFLLTLWTOVERKIQKDSLKPKLGTGTAGOMCSGLPLSSV 484
QY 547 BENKTCESNNTS-----ATTTSIQPNL--ENSNSSELNSSQSESAAA-----588
Db 485 DINNFGDSINSEGIPLKRRRVSGFCHLRPELFDENLPNTPLRGEAPTNRKSLVMHTP 544
QY 589 -----DDPE--NGERES-----HTPVSIQEEIVG-----DFTSEKSTGELSESPG 626
Db 545 PVLKKIIEQPOPQSGKQSGSEIHVEVKAQSLVISPPAPSPRKTVPASDQRRRSCKTAPA 604
QY 627 AGKGAS-----GSTRILITRLNPDS-----KLSQLKSQVAAAHAEANKLFKRGKVL 674
Db 605 SSSKSQTEVPKRGGERVATCLOKRVVISRSOHDILQICSKRRSGASEAN-----LI 656
QY 675 VVNSQGEISRLSTKK---EVIKMG---NINNYFKLGQEGKYRV--YHNOYSTN-----S 720
Db 657 VAKSWADVVLKAKQOTQVIKHGQORSNMNRQRPATPKPVGVEHVSQFSTGHANSFCT 716
QY 721 FALNKQHUREDHKRR--HLAHKFCLTPAGFEK-----WNSVHGSKVLTISTLRL 769
Db 717 IIGRAHTEKVVHVPARPYRVNLNFTISNOKMDFKEDLSGIAEMFKTPVKEQPOLT--STCHI 775
QY 770 TITOLENIPSSF-----FHPNASHRANWIKAVQCMCK--PREFALALAILLECAVK 819
Db 776 AINSNENLGGKFOQTDSGEPLLTSESFGNVEFFSAQNAKQPSD-----KCSAS 827
QY 820 PYVMPLPIWREFLGHT-----RLHRMTSIERE-----EKEK 849
Db 828 P-----PLRRQCIRENGNVAKTPRNTYKMTSLTKTSDTEPESKTVSTVNRSGRSTEFNR 883
QY 850 VKK--KEKKQBEETM-----QOATWVKYTFPVKHQVHWKQGEYRYVGYGGSWVI 898
Db 884 IQKLPVESKEETNTEIVECILKRGOKATLQO-----RRGEMKEI-----925
QY 899 SKTHVYRVFKLPNGNTNVYRKSLEGTKNNMDENMDESK-----RKCS--RSPKKIKIE 951
Db 926 -----ERPFET--YKENIELKEN--DEKMKAMKRSTWQOKCAPMSDLDLKL 970
QY 952 PDSE--KDEVKG-----SDAAKAGADQNMEDISKITEKKOODVKELLDSDSKPCKE-EP 1002
Db 971 POTELEMDTARGONLLQTDHAKAPKSEKGIKTM-----PCQSLOP 1012
QY 1003 MEVDDDMKTESHVNCQESSQVDVNVSE-----GHLRTS-----Y 1038
Db 1013 ----EPINTPTTKOOLKASLGKVGKVELLAVGKFTTSTGETTHTHREPAGDGKSIRTF 1068
QY 1039 KKKTK-----SKLDGLLERRITKQFTLEKORLEKIKLEGGIKGIGKTSNKSNLSESP 1093
Db 1069 KESPKOILDPAARVTGM--KKWPRTPKEAQASLEDL---AGFKELFQTPGSPSESMTEK 1123
QY 1094 VITKAKGQCSMSRMQEOSPNANDQPE--DLIQGCSQSDSSVLRMSDPSTHTNKLYPKDR 1152
Db 1124 T---TKIACKSPPPESVDPTSTKQWPKRSLRKADVEEFLALRLKLTPSAGKAMLTPKPA 1180
QY 1153 VLDDVSI-----RSPETKCPKQNSIEN-----DIEKVSUL 1183
Db 1181 GGDEKDIAFMCTPVOKLDLAGLTPGSKROLQTPKEAQALEDLAGFKELFOTPGHTEEL 1240
QY 1184 ASRGO-----EPTKSKTKGNDFIDDSKILASADDITGLT--CNKKPLQIE 1227
Db 1241 VAAGTKTKIPCDSPQSDPVDVTSTKQR-----PKRSIRKADVEGELLACRNLM-----1290

QY 1228 ESDTIVSSSKSALHSSVPKSTNDRD-----ATPLSRAMDFEGKLGCDSESNSTLENSDST 1282
Db 1291 -----SAGKAMHTPKPSVGEKNDIIFVGTVPQKL-----DLTENLTGS 1329
QY 1283 VSIQDSSEEDMIVONSNESISEQFRTREODQVEVLEPLKCELVSGESTG--NCEDRLRPVKGT 1341
Db 1330 KRRPQTPKEEAQALEDLTGFKELFQTPGHTEEA-----VAAKTTKMPCESPPPESAD 1382
QY 1342 EANGKKPSQOKKLEERPVNKCSDOIKLKNTTDDKKNNENRESEKKGQRTSTFOINGKONKP 1401
Db 1383 TPTSTRQPKTPLEKRDVQKELSALK-----KLTQTSGETHTHDKVPGGEDKS 1430
QY 1402 KIYLGECLEKISESRVVSNGV--EPKVNINIKIIPENDIKSLTYSKSAIRFINGDVIME 1460
Db 1431 INAFRETAQOKLDPAASVTSKSRHPKTK--EKAQPLEDLAGW--KELFQTP-----VCT 1480
QY 1461 DFNERNSETSKHLLSSSDAEGNRYDSLETLPSTKESDSTQTTTPSASCPSNVSQVED 1520
Db 1481 DKPTTHEKTKIACRSQPD-----PVDTPTSKPSKSLRKVD 1519
QY 1521 MEIETSEVKKYTSS-----PITSEESNLNDFIDENGLPINK-----NENYNGESKR 1568
Db 1520 VEEFFALRKRTTPSAGKAMHTPKPAVSGEKNIYA--FM--GTPVQKLDLTLENLTGSKRR 1574
QY 1569 KTVITEVTMTSTVA-----TESTVTIKVEKGDKQTVVVSSTENCAKSTVTTT 1616
Db 1575 LQTPKEAQALEDLAGFKELFQTRGHTESMT--NDKTAKVACKSQSDPLDKNPASSKR 1631
QY 1617 TTVTKLSTPSTGGSDIISVKEQSKTVVTTVTDLSLTTGTGLTVMTSKEYSTRDKVK 1676
Db 1632 RLKTSLG--KVGVEKELLAVGLKTQTSGETTHTHTPTGDKGSKMAFNPESKQIILDSAS 1689
QY 1677 LMKFSRPKTRSGTA-----LPSYRK--FVTKSTKKSIFVLNDDOL-----1715
Db 1690 LTGSKROLRTPKGRSEVPEDLAGFIELFQTPESHKESMTNEKTTKVSYRASOPDLVDNPT 1749
QY 1716 -----KKLARKGGIREVPYFNNAKPALDIWPPSPRPTFGITWRYLQIVKSLAGVSL 1769
Db 1750 SSKPQKRSLRKADTEE-----EFLAFRKQTPSAG-----KAMHTPKPAVGEEK 1793
QY 1770 MLR--LLWASLWDDMAAKVPPGGSTRTTETETETITTEILIKRRDVGYPGIR-----1820
Db 1794 DINTELGTPVQKLDQGNLP--GSNRLOTRKEKAQALELT-----GFRELQTPCT 1844
QY 1821 -----FEYCIRKIIC--PIGVP--EPPKETPTPQRKGLRSS-----ALRKP--1859
Db 1845 DNPTTDEKTKKILCKSPQSDPADPTNTKQRPKRSLLKADVEEFLAFLRKLTPSAGKAM 1904
QY 1860 ETPKQTPGVITETWVAEELELEWEIRAF-----AERVE-----KEKAQAVE 1900
Db 1905 HTPKAA-----VGECK-----DINTFVGTPVEKLDLLGNLPGSKRRPQTPKEKAKALE 1952
QY 1901 QOAKRLEQOKPTVIATSTTSSTTSTISPAQKVMVAPISGVSVTGTTKMWLTTKVGPSP 1960
Db 1953 DLAGFKELFQTPGHTHEESMTDDKITEVSKSPQDPVKTP-----TSSKQRLKLSLCK--2005
QY 1961 ATVTFOQKNFHTPATVWKQGSNGVVQOQKVL--GIIPSSGTGTSQOFTTSFQPTA 2018
Db 2006 -----VGVEEVLVPVKLTQTSKGTQOT-----2028
QY 2019 TVTIRPNTSGSGT-----TSNSQVI-----TGPOIRPGMTVIRTP-----LQOSTLG 2061
Db 2029 -----HRETAGDGKSTKAFKESAKQMLDPANYGTGMRWP-----RTPKEEAQSLDLAGF 2079
QY 2062 KAIRITVWVOPGAQPOQVMTQIIRQOPVSTVSAPTVSSTPGQKSLTSATSTSIQSSA 2121
Db 2080 KELFOTPDHTESTTDDKTKI-----ACKSPPPESMDTPTST-----2117
QY 2122 SOPPPRQOQVKLT--MAQLTOLTQGHGNGOGLTVVIOGOGTQTLQOLIP-----QGVTV 2175
Db 2118 RRRPKTPLGKRDIIVEELSALKOLTQ-----THTTDKVPGEDEKGINV 2159

```
QY 2176 LPPGGOQLM-----QAAMPNGTVQ-----RFLF-TPLATATTATTSTTTTV 2215
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2160 FRETAKQKLDPAASVTSKSRQPTPKGAQKPLEDLAGLKFQTPCTDKPTTHEKTKI 2219
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2216 STTA-----AGTGE--QROSKLS-POMOVHODKTLPPAOSSSVSGPAKAOPOTA----- 2260
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2220 ACSRPQDPVGTTFIFKPSKSLRKADVEESLALRKRTSPVSGKAMDTPKPAAGDEKDM 2279
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2261 -----OPSARPOQPTQPSPAQPEVQTOPEVQTOQTIVSS-----HYVPSAQPT 2303
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2280 KAFMGTPVQKLDLPGLNPGSKRPQTPKE-----KAQALEDLAGLKFQTPGTDKPT 2332
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2304 HAOSSPQVAQAQSQPOSNOGQSPVRVQSPQTRIRPS-----TPS- 2344
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2333 -TDEKTKIAKKS-PQ-----PDFVDTPTASTKQRPKRLRKADVEEFLALRKRTPSA 2383
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2345 -----QLSPGOQSQOVQTTSTQPIOPHTSL-OIP-SOGOPOQOSOP----- 2382
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2384 GKAMDTPKPAVSDEKKNINFVETPVOKLDDLGNLPGSKRQPTPKKEKALEDLVGFKEK 2443
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2383 -QVQSTQTLSSGQTLNOVSVSPRPOLQIQOPQVIAVP-----QLQOOVOVLQSI-- 2435
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2444 FQTPGHTESMTDDKITEVSCSKSPQSPESFKTSRSSKQRLKIPLVKVDKKEEPLAVSKLTR 2503
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2436 OSQVVAQIQAOQSGVQOIKL--OLPIQOOSAVOTHQIQNVTVVQAASVQEQLOQVQ 2493
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2504 TSGETTQTHTEPTGDSKSIKAFKESPKILDPAA-----SVTGSRRQLRTRKEKARA 2555
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2494 LRDOQOKKK-----QOOIEINVN-----TPSKLLIKVELIQOVVM 2529
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2556 LEDLVDFKELFSAPGHITESMTIDKNTKIPCKSPPPPELDTATSTKRCPKTRL--RREVK 2613
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2530 KHNVAIEHLKQKKSMT-----PAREENQRMIVCNQVMYILDKIDKEEKQAARKRE 2583
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2614 EELSAVERLTQTSQSTHTHKEPASGDEGKVL-----KQRAKKKPNP 2656
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2584 ESVEQKRSKQNA-----TKLSALLFKHKEQLRA-----EILKK 2616
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2657 VEEEPSSRRPRAPKEKAQPLEDLAGFTLSETSGHTQESITAGKATKIPCESPPLEVDT 2716
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2617 RALLDLDQTEVOEELKRLDIKKERDLMOLAQ-----AT 2651
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2717 TASTKHLRTRVOK-----VOVKEPSAVKFTQTSGETTDADKEPAGEDKGIKALKESAK 2771
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2652 AVAAPCPVTVPLPAPAPPSP-----PPPGVQHTGLLSTPLPVASOKRKEE 2703
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2772 QTPAPAAVTSGRRRPRAPRESAQATJEDLAGFKDPAAGHTEESMTDDKTKIPCKSSPEL 2831
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2704 KDSKSKSKK-----KWIISTSKETKDKTKLYCICKTPYDESK 2741
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2832 EDTATSSKRRPRTRAKQVEVKEELLAVGKLTQTSGETTHTD-----REPVGEGK 2880
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 6

```
US-09-919-039-21
; Sequence 21, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919, 039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 21
; LENGTH: 3256
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

```
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2700132CD1
US-09-919-039-21
```

```
Query Match      2.9%   Score 438;   DB 11;   Length 3256;
Best Local Similarity 18.3%;   Pred No. 1.9e-13;
Matches 560;   Conservative 418;   Mismatches 1068;   Indels 1010;   Gaps 143;
```

```
QY 300 PVPQVTCVAFIQKNKPYIRHEPIGYDRSRKRYWFLNRLRIIEEDTENENEKKIWIY----- 354
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 221 VP--TTQCLDMSKNE-----SPFWKLYESVKKLELDVKSQKENVLYQYCRKSG 265
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 355 -----YSTKVQLAELIDCLDKDYWEAECLKILEEMREBIHRHMDITEDLTWKARGSNKSF 410
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 LQTDYATEKESADGLQ-----GETQLL-----VSRKSRPKSGSGHAVA 304
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 411 AAAN-EEILESIRAKKGDIDNVKSPET-----EKDKNET-----ENDSKDAEKNR 455
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 305 EPASPEQELDONKGGKRDVESVQTPSKAVGASCFLEYEPAMKMTPVQVSOQONSFPQKHNK 364
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 456 EEF-----EDQSLKED-----SDDKTPDDPDEQKSEPTVEVGDKGNSVS--ANLGDNTTN 504
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 365 DLYTTGRRESVNLGKSEGFKAGDKTLTPRKLTSTNRNTPAKVEDAADSAATKPNLSKTRG 424
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 505 A--TSEETSPSE--GRSPVGCLET-----PDSSNMAEK-----KVASELPQDVPE 546
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 SIPTDVEVLPTETEIHNNEPFLTLWLTOVERKIQKDSLKPKEKLTGTAGQMSCGLPLGSSV 484
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 547 EPNKTCESNNTS-----ATTTTSIQPNL--ENSSSELNSSQESAKAA----- 588
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 485 DINFGDSINESGIDPLKRRRVFGHLRPELFDENLPPNTPLRKGAPTKRSLVHMTP 544
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 589 -----DDPE-NGERES-----HTPVSIQEEIVG-----DFTSEKSTGELSSEPG 626
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 545 PVLKKIIEQOPQSGKQESGEIHVEVKAQSLVISPAPSPRKTVPASDQRRRSCKTAPA 604
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 627 AGKAS-----CSTRIITRLNPD-----KLSQLKSOQVAAAAHEANKLFKEGKVL 674
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 605 SSSKSQTEVPKRGGERVATCLOKRVISRSQHDILQIMICSKRRSGASEAN-----LI 656
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 675 VVNSQGEISRLSTKK--EVIMKG--NINNYFKLQOEGKYRV--YHNOYSTN-----S 720
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 657 VAKSWADVVKLGAKQOTQVKVKGQPSRMKRRORRATPKPVGEVHSQFSTGHANSPT 716
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 721 FALNKHOHREDHDKRR--HLAHKECLTIPAGEFK-----WNGSVHSGKVLITISPLRL 769
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 717 IITGKAHTEKVHVPARPYRVLNNFISNQKDFEDLSGIAEMFKTPVKEQPLT-STCHI 775
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 770 TITOLENNIPSPF-----PHPNWASHRANWIKAVQCMCK-PREFALALALECAVK 819
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 776 AINSNENLLGKQFGTDSGEEPLPTSESGGNVFFSAQNAAKQPSD-----KCSAS 827
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 820 PVVMLPTWREFLIGHT-----RLHRMTSIERE-----EKEK 849
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 828 P-----PLRROCIRENGNVAKTPRNTYKMTSLETKTSDTEPSTKTVTVNRSRSTFRN 883
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 850 VKK--KEKKOEBEETM-----QOATWVKVTFPVKHQVHKOKGEEYRVTVGGWSWI 898
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 884 IQKLPVESKSEETNTEIVECILKRGOKATLLQ-----RREGEMKEI----- 925
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 899 SKTHVYRFVPKLPNTNVNRYKSLGKTKNMNDENMDESCK-----RKCS--RSPKKIKIE 951
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 926 -----ERPFT--YKENIELKEN--DEKMKAMKRSRTWQKCAPMSDLTDLKSL 970
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 952 PDSE--KDEVKG-----SDAAKGADQNEMDISKITEKKDQDVKELSDSDPKCKE-EP 1002
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 971 PDELMKDARGONLLQTDHAKAPKSEKGIKTM-----PCQSLOP 1012
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1003 MEVDDDMKTESHYNCQESSQOVVVNVSE-----GFHLRTS-----Y 1038
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1013 -----EPINTPTHTKQOLKASLGKVGKVELLAVGKFTRTSGETHTHREPAGDGKSIPTF 1068
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1039 KKKTK-----SSKLDGLLERIKQFTLEEKQRIKLEGGIKGICKTSTNSKNLSESP 1093
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

1069	Db	KESPQILDPAAKVGM--KKWPRTPKEEAQSLEDL---AGFKELFQTPGCPSESMTEK	1121
1094	Qy	VITKAKEGQSDSMROEQSPNANDPOE-DLIQCQSQSDSSVLRLMSDPSHTNKLYPKDR	1152
1124	Db	T---TKIACKSPPPESVDTPSTKQWPKRSLRKADVEEFLALRLKLTSPSAGKAMLTPKPA	1180
1153	Qy	VLDVSI-----RSPETKCPKQNSIEN-----DIEEKVSDL	1183
1181	Db	GSDEKIDKAFMGTPVQKLDLAGTLPGSKKQLQTPKEKAQALEDLAGFKELFQTPGHTEEL	1240
1184	Qy	ASRGQ-----EPTKSKTKGNDFFIDDSKLASADDITGLI-CKNKKPLQIE	1227
1241	Db	VAAGKTKIPCDSPQSDPVDTPSTKQR-----PKRSIRKADVEGELLACRNLP----	1290
1228	Qy	ESDTIVSSKSAIHVSVPKSTNDRD-----ATPLSRAMDFEGKLGCDSESNSTLENSDST	1282
1291	Db	-----SAGKAMHTPKPSVGEEDKIIIFVGTGPVQKL-----DUTENLTGS	1329
1283	Qy	VSIODSEEDMIYVONSNEISIEQFRREQDVEVLEPLKCLVLSGESTG-NCEDRLPVKGT	1341
1330	Db	KRPQTPKEAQALEDLTGFKELFQTPGHTEEA-----VAAGKTKMPCESPSPESAD	1382
1342	Qy	EANGKPKSQOKKLEERPYNKCSQOIKLNTTDRKKNNENRESEKKGORTSTFQINGKDNKP	1401
1383	Db	TPSTSTRQPKTPLEKRDVQKELSALK-----KLTQTSGETHTHDKVPGEDKS	1430
1402	Qy	KIYLGECLKEISESRVVSQNV--EPKVNININKIIPENDIKSLIVKESAIRPFINGDVIME	1460
1431	Db	INAFRETAQKLDPAASVTSKGRHPKTK--EKAQPLEDLAGW--KELFQTP-----VCT	1480
1461	Qy	DFERNSESFTKSHLLSSSDAEGNVRDSLETLPSTKESDSTQTTTPSASCPSNSVQVED	1520
1481	Db	DKPTTHEKTKIACRSQPD-----PVDTPTSKPKQSRSLRKVD	1519
1521	Qy	MEIETSEVKKVTS-----PITSEESNISDNFIDENGLPINK---NENNVNSESKR	1568
1520	Db	VEEFTFALRRTPSAGKAMHTPKPAVSGEKNIYA--FM--GTPVQKLDITENLTGSKRR	1574
1569	Qy	KTVITEVTTMTSTA-----TESKTVIKVEKGBQKQVWSVSNCAKSTVTTT	1616
1575	Db	LQTPKEKAQALEDLAGFKELFQTPRGHTEESMT--NDKTAIVACKSQPDLDKNPASSKR	1631
1617	Qy	TTWTKLSTPGSGVDIISVKEQSKTVVTTVTDLSLTTGGTLVTSMTVSKVEYSTRDKVK	1676
1632	Db	RLKTSLG--KVGVEKEELLAVGLKLTQSGETHHTHTPTGDKGNKAFMESPQKILDSAA	1689
1677	Qy	LMKFSRPKKTSGTA-----LPSVRK--FVTKSTNKKSTFVLPNDL-----	1715
1690	Db	LTSKRLRTPKSGSEVPEDLAGFIELFQTPSHTKESMTNEKTKVSYRASQDPLVDTP	1749
1716	Qy	-----KKLARKGGIREVPVFNNAKPALDIWYPSPRPFGITWRYRLQTVSKLAGVSL	1769
1750	Db	SSKPPQKRSLRKADTEE-----EFLAFRKQTPSAG---KAMHTPKPAVGEEK	1793
1770	Qy	MLR--LLWASLEWDDMAKVPGGSGSTRTETSEVITTEILIKRDRVCPYGR-----	1820
1794	Db	DINTFLGTPVOKLDQPCGNLP--GSNRLOTRKEKAQALEET-----GFRELFQTPCT	1844
1821	Qy	-----FVYCIKRIIC--PIGVP-ETPKETPTPORKGLRSS-----ALRPKR-----	1859
1845	Db	DNPTTDEKTKKILCKSPQSDPADTPNTKQPKRSLKKADVEEFLAFKRLTPSACKAM	1904
1860	Qy	ETPKQTPGVIIETWABEELEHWRAP-----AERVE-----KEKAQAVE	1900-
1905	Db	HTPKAA-----VGEK-----DINTFVGTPEVKELDLLGNLPGSKRRPQTPKEKAKALE	1952
1901	Qy	QOAKKRLQOKPTVIATSTTSPTSSTTSPACKVMVAPISGSVTTGTVMVLTTKVGVSP	1960
1953	Db	DLAGFKELFQTPGHTEESMTDDKITEVSCSKSPQDPVKTP-----TSSKQRLKISLGK-	2005
1961	Qy	ATVTFQONKNFHQFATVWVKOGOSNGVVOVQOKVL--GIIPSTGTSGTOOTFTSFQDPTA	2018
2006	Db	-----VGVEEVLVPVKLTQTSKGTOT-----	2028

Qy	2019	TVTRPNYSSGGT-----TNSQVI-----TCPQIRPGMVTIRTP-----LQOSTLG	206
Db	2029	-----HRETAGDKSIKAFKESAKQMLDPANYGTMERWP-----RTPEAKQSLEDIAGF	2079
Qy	2062	KAIIRTPVMQPGAPQOQVMTQILIRGQPVSTAVSAPNTVSSTPGOKSLTSTATSNIQSSA	2121
Db	2080	KELPQTPDHTEESTDDKTKI-----ACKSPPPESMDTPST-----	2117
Qy	2122	SQPRPQOQGVKLA--MAQLTQLTQGHGNGOGLTVVIQOGQGTGLOLIP-----QGVTV	2175
Db	2118	RRREKTPLGKREDIVEELSALKQLTQ-----TTHTDKVPGEDDKGIN	2159
Qy	2176	LPFGGQOLM-----QAAMPNGTVQ-----RFLF--TPLATTATTATTTTTV	2215
Db	2160	FRETAQKLDPAASVGTGSKRPRTPKGAQPLEDLAGLKELFOPTICTDKPRTHEKTTKI	2219
Qy	2216	STTA-----ACTGE--QROSKLS--POMQVHODKTLPPAOSSSVGPAPQAQOTA-----	2260
Db	2220	ACRSPQDPVGTPTIFPKQSKRSRKADVEESIALRKRTPSVGKAMDTPKAGGDEKDM	2279
Qy	2261	-----QPSARQPQOTQSPQAQPEVQTQPEVQTOITVSS-----HVPSEAQPT	2303
Db	2280	KAFWGTVPQKLDLPGNLPGSKRWQTPKE-----KAQALEDLAGFKELFQTPGTDKPT	2332
Qy	2304	HAQSKPKQAAQSQPQSVNGQSPRVOSPSQTRIPS-----TPS--	2344
Db	2333	-TDEKTTKIACKS-PQ-----PDVDTPASTRQRPKNRKADVEEFLALRKRTPSA	2383
Qy	2345	-----QLSPGQOSQVQITTSQPIQPHTSL-QIP-SQOGQSQSQP-----	2382
Db	2384	GKAMDTPKPAVSDKNIINTFVETPVQKLDLGNLPGSKRQPTQPEKAEALDLVGKEL	2433
Qy	2383	-QVQSSTOTLSSGOTNLNVSVPSSRPLOLQIQQPQVIAVP-----QLQQQOVQLSQI--	2435
Db	2444	QTPGHTHEESMTDKIIEVSCCKSQPESFKTSRSKQRLKPLVKVDMKEEPLAVSKLTR	2503
Qy	2436	QSQVVAQIAQOOSGVPOQIKL--QLPIQIOSSAVQTHQIONVTVQNAASVOEQLQRVQO	2493
Db	2504	TSGETTQTHTEPTGDSKSIKAFKESPKOILDPAA-----SVTGSRRQLRTRKEARA	2555
Qy	2494	LRDQOQKKK-----QQQIEINVN-----TPSKLLIKVEIIQOKVVM	2529
Db	2556	LEDLVDFKELFSAPGHTHEESMTIDKNTWKIPCKSPPELTDATSTKRCPKTRL--RKEVK	2613
Qy	2530	KHNAVIEHLKOKKSMT-----PABREENQRMIVCNQVMYKILDKIDKEEQAAKKRRE	2583
Db	2614	EELSVERLTTQSGOETHKEPASGDEGKIVL-----KQAKKKKPNP	2650
Qy	2584	ESVEOKRSKQNA-----TKLSALLPKHKQLRA-----EILK	2616
Db	2657	VEEESRRRRPRAPKEKAQPLEDLAGFTELSETSGHTQESLTAGKATKIPCESPPELVVD	2716
Qy	2617	RALLDKDLQIEVQBELKDLKIKKEKDLMLAQ-----AT	2651
Db	2717	TASTKRHLRTRVQK-----VQVKEPSAVKTTQTSGETTDADKEPAGEDGKIKALKESAK	2771
Qy	2652	AVAAPCPVTPVLPAPPAPPPSP-----PPPGVQHTGILLSTPTTLFVASQKRKREE	2703
Db	2772	QTPAPAAVYTSRRRPRAPRESAQAIEDLAGFKDPAAGHTHEESMTDDKTTKIPCKSSPEL	2833
Qy	2704	KDSSSKSKKK-----KMISTTSKETKDKTKLYCICKCTPYDESK	2741
Db	2832	EDTATSSRRRPRTRAKQVEKVEELIAGKLTQTSGETTHD-----KEPVGECK	2880

RESULT 7
US-09-864-761-34248
; Sequence 34248, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.

Db 1127 QAESNVDPEDPSTQPLSRPAQKSEANEPAEKDPDADAEPAQNAQKAAAPESQPPASE 1186
QY 1020 SSQVDVNVSEGFHLRTSYKKKT-----KSKLDGLLERIKQFTLE-- 1061
Db 1187 DLEVDPVPAKDKPKNKRKSKTPYQAAAVSIVEKPVTRKSERID---REKLKRSNPRG 1243
QY 1062 EKQREKIKLEGKIGIKTSTNSKNLSESPVITKAKEGQCSMRQEQSPNANNDQPE 1121
Db 1244 EAQKLELAKMEA--EKIIRTAASKNAADLEHPEPSLPSLRRRNRVRSYVATMGDHENS 1301
QY 1122 DLIQCSQSDSVLRMSDPSTHTNKLKPKRDVLDDVSI-----RSPETKCPKQNSIE 1173
Db 1302 PVKEPVEQ-----PRVTRKRL---ERELQEAATAVPTPRGRPPKTRRRADEEEE 1348
QY 1174 NDIEKVSDL--ASRGQPTKSKT-----KG-NDFFIDDSKLASADDIGTLICKNK 1221
Db 1349 NEAKEPAETLKPPEGWRSRQSKTAAGGPGQKGNKPEKVDATREATTEVGPQIGYKE 1408
QY 1222 KPLIOESDTIVSSKSKALHSSVPKSTNDRDATPLSRAMDPEKGLGCDSESNSTLENSD 1281
Db 1409 SWEPKAAEEAGSQKDRDKDAGTDKNPPETAPVEVV---EKKPAPEKNSKSKGRSRN 1465
QY 1282 TVSIQDSSSEDMIVQNSNIS-----EQFRTREQDVEVLEPLKCELVSGESTGNCEDRL 1336
Db 1466 SRLAVDKSAS---LKNVDAVSPRGAAGAAQAGERESGVAVSPEKSE----- 1508
QY 1337 PVKGTENGKKPSQOKLEERPVKCSQDIKLNKNTDKNNENRESEKKGORTSTFOING 1396
Db 1509 -----SPOKE-----DGLSOLKSDPVDPKPEKEDVSAGSPSPATQL-- 1548
QY 1397 KONKPKIYLGBCLEISESRVSVGNVEPKVNNINKIIPENDIK-----SLTVKESAIRP 1451
Db 1549 ----AKOMELEQAVEHIAKLAESASAAKYADAPGLAPEDRDKPAHQASETELAAGS 1604
QY 1452 FINGVIMEDFN-----PRNSETKSHLLSSDARG 1482
Db 1605 IIN-DISGEPENFPAPPYPGESQTDLOPPAGAQAALPSEEGMETDEAVSGILETEAATE 1663
QY 1483 NYRDSLETL-PS-----TKESDSTOTTPSASCPSNSNOVEDMEIETSEVKKTSPI 1536
Db 1664 SRPPVNAADPPAGPTDTEARGNSSET--SHSVPEAKGSKEVEVTLVRKDKGRQKT--- 1719
QY 1537 TSEESNLNDFIDENGLPINKNENVNGESKRTVITEVTTMTSTVATESKTVIKVEGD 1596
Db 1720 RSRKRNTNKKVVAVPESHVPESNOAQGES---PAANEGETTVQHPPEAPOE-----EKOS 1770
QY 1597 KQTVSVSTENCASVTITTTVTVKLSTPSTGSGVDIISVKEQSKTVVTTTVDLSITTTG 1656
Db 1771 EKPHSTPPQSC-----TSDSLKI--PSTENSSQEIISVEERTPT--KASVPPDLPPP- 1817
QY 1657 GTLVTSMTYSKEYSTRDKVYKLMKFRPKKTRGTALPSYRKFKVTKSKSIFVLNPDLLK 1716
Db 1818 ---POPAPVDEEPOAFRHHIIESDPVTPSPDSIP-----IPTLPSVTA 1861
QY 1717 KLA---RKGIREVPYFNNAKPALDIWYPSPRPTFGITWRYRLQTVKSLAGVSLMLKL 1773
Db 1862 KLSPPVASGGI-----PHQSP-PTKVTEWITRQEPRAQSTPS----- 1898
QY 1774 LWASLRWDDMAAKVPPGGGSTR-----TETSETEITTEILKRRDVGPGYIRFEYCIRK 1827
Db 1899 --PALPPDTKASDVDTSSSTFLRKILMDPKYVSATSVTSTSV-----TTA 1940
QY 1828 IICPIGVETPKETPTPQRKGLRSSALRPKRPETPKQTFGIVILETWVAEELELWEIRAF 1887
Db 1941 IAEPVSAAPCLHEAPP-----PVDSSKKPLEEKTAPPV-----TNNSEIOASEVLVA 1987
QY 1888 AERVEKEKAQV-----EQQAKRLEQCKQPTVIATSTTSSTSTST 1929
Db 1988 A---DKEKAPVAPKITSIVSRMPVSIDLENSQKTLAKAPQOTLGLVSALTGLVNVVS 2044
QY 1930 ISPAQKVMVAPISGVTGKWLTKVGSFAPVTTFOQKNFHTQFATVWVKQGSN--SG 1987

Db 2045 LVPV-NALGPVKQSVTT-LKSLVSTPAG-PVNVL-----KGPVNVLTG 2085
QY 1988 WVQV-----QOKVLGIIPSTGTSTQOTFTSFQPTATVITRPN--TSGSGGTTNSQVIT- 2040
Db 2086 PVNVLTTPVNATVGTVNAACPCTVNAASAVNATASAVTGTAGAVTAASGGVATTGTGVTM 2145
QY 2041 -GPOIRGEMTVIRTPLOQSTLGLKAIIRTPVMVQPCAPQOVMTQIIRQOPVSTAVSAPNTV 2099
Db 2146 AGAVIAPS-----TKCKQ-----RASANENSFRHFGS-----MPVIDDRPADAGSGAGLRV 2191
QY 2100 SSTFGQSKSLTSATSTSNIOSSASOPPRPOQOVKLTMAQLTQLTQGHGQGNQLTVVITQGG 2159
Db 2192 NTSQGVLLS-----YSGQTEGPQR-----ISAKISQIIPA-----SAMDIEFQ-- 2231
QY 2160 QOTTGQQLQILPQGVTVL--PGPGQQLMQAAMPNGTVORFELFTPLATTATTASTTTT---- 2213
Db 2232 -QSVYSKQVKPDSVTASQPPSKGPQ-----APAGYANVATHSTLVLTQAQTYNASPVISV 2285
QY 2214 -----TVST--TAAGTGE-QROSKLSQMQVHQDKTLPPA-----QSSSVG 2251
Db 2286 KADRPSELEKPEPIHLSSVSTPVTOGGTVKVLTOGINTPPVLVHNLQVLTPPSIVTTNKKLAD 2345
QY 2252 PA--KAOPQTAQPSARPQOPQOPQAPQ-----EVOTQPEVOTQTTVSSHVPSEAQ 2301
Db 2346 PVTUKIETKVLQPMGLSTLPHHPALPSKLPTVEVNHVPSPGIPADRTV-SHL--AAA 2402
QY 2302 PTHAQSSKPOVAAQSQOSQVQSPVRVQSPQSPQSPQSPQSPQSPQSPQSPQSPQSPQSPQSP 2361
Db 2403 KLDHSPRP-----SGPCPSFPRAS-----HPSSTASTALSTNATVLAAGIP 2446
QY 2362 IP-----IQPHTSLQIPSQGPQ-----SQPOVQSSTOTLSS-GOTLNQVSVSSPSRP 2408
Db 2447 VPQFTSSIHPEQSVIMPHPSITQVLSHLSQGEVRMNTPTLPSITYSIRPEALHSPAP 2506
QY 2409 OLOIQQ-----PQPOVIAVQLOQQ-----VOVLSTQOSQV--A 2441
Db 2507 -LQPOQIEVRAPORASTQAPAGVPALASQHPPEEVHYHLPVARATAPQOSEVLVQMS 2565
QY 2442 QIAQOOSGVQOIKQLPQIQOOSAVQTHQIQNVTVVQAAASVQEQ 2487
Db 2566 EYRLHPYVTPRDRVIMVPHVHTAVSE-QPRAADGVKVPVPKAPQ 2610

RESULT 8

US-09-738-973-425
; Sequence 425, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738, 973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 425
; LENGTH: 4019
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-425

Query Match		2.7%	Score 402;	DB 10;	Length 4019;
Best Local Similarity		17.3%;	Pred. No. 1.8e-11;		
Matches 510;		Conservative 314;	Mismatches 992;	Indels 1126;	Gaps 108;
QY	254	CRVCHKL---	GDLCCETCSAVYHLECVKVPLEEVPEDEMOCEVCAHVKGVTDCVAEI	310	
Db	65	CEAGKATDPRLLJLCCDDCDSYHYCLDPLQTPVKGWKCKWCVCRHCGATSAGLRC	124		
QY	311	QKNPYIRHEPIGYDRSRKRYWFLNRLIIEEDTENENENKKIWIYSTKVQLAELIDCLDK	370		
Db	125	EWONNYTQCAPCA-SLSSCPVCYRNYR---	EEDL-----	ILQCRQC	161
QY	371	DYWEAECLKILEEMREEIHRMDITEDLTNKRAGNSKSFLLAANEELIESIRAKKGIDIN	430		
Db	162	DRMHAVQCNL-NTEEEVENVADIGFDCS-----	MCRPYMPAS-----	N	199
QY	431	VKSPEETKDKNETENDSKOAEKNREEFEDQSLKSDDKTPDDDPQGGKSEETEVGDK	490		
Db	200	VPSDCEB-----	SSLVAQIVTKVELOPPKTYTQGV-----	232	
QY	491	GNYSANLGDNTTATSEETSPSEGRSPVGCSETPDSSNMAE-KKVASLELPQDVPPEPN	549		
Db	233	-----	CLTE-----	SGMTQLSLTVTVPRKRKPK	258
QY	550	KTCSSN-TSATTTSIQPNLENSN-SSELNSSQSESAKAADDPENGESHTTPVSIQEE	607		
Db	259	LKLKIINONSVAVLQTPDIOSEHSRDCGDDSGELMDCD-----	GKSES-----	SPERE	310
QY	608	IVGFTSEKSGELSESPGAGKAGSGSTRIITRLNRNPDLSQLSKSQVAAAHEANKLF	667		
Db	311	AVDEET-----	KGVEGTDGVKKRKKP-----	Y	333
QY	668	KEGKVLVYNSQGEISRLSTRKEVIMK-----	GNINNYFKLGOEGYRYVYHNOYSTNSFALN	724	
Db	334	RPGIGGEMVQRSGTGKTKRSVIRKDDSSGSISEQLPCRDDG-----	376		
QY	725	KHQHREDHDKRRHLAHKFCLTYPAGEFKWNGSVHGSKVLTTISLTLLTTQLENNIPSPFFH	784		
Db	377	-----	376		
QY	785	PNWASHRANWIKAVQCMCKPREFALALALECAVKKVVMVLPPIWREELGHTLRHMTSIE	844		
Db	377	-----	WSEQLPDTLVDSESVV-T	393	
QY	845	EKEKVKKKKKQBE--EET----	MQATWVKYTFPVKQVWKQKGEEYRTVYGGWSWI	898	
Db	394	ESTEKIKRKRKRKNKLEETFPAYLQEAFFCKDLDDTSRQ-----	SKI	436	
QY	899	SKTHYRVVPKPGNTNVNRYKSLGTYKNNMDENDESDDRKCRSPKKIKIEPDSEKDE	958		
Db	437	SLDNLSEDAQLLYKTNNM-----	TGFLDPSLD--	PLLSSSSAPT-K	475
QY	959	VKGSAAKAGADONEMDISKITEKKDQDKVELLSDSDKPKCEEPM-EVDDDMKTESHVNC	1017		
Db	476	---SGTHGPADPLADISEVL-NTDDDLGLIISDLAKSVHDHSDIGPVTDPSLPQPNV	531		
QY	1018	QESSQDVVNVVSEGFHLRTSYKKTKSKLDGLLERRIKQFTLEKORLEKIKLEGGIGK	1077		
Db	532	NOSS-----	RPLSEQLDGLI-----	SPELDKMVTDGALG	562
QY	1078	-IGTSTNNSKNLSESPVITKAKGCGSDSMRQEGSP-----	NANNQDEPDLIQGCSQ	1129	
Db	563	KLYKRIPELGGKDVEDLEFATVLSPANTQPTPLPQPPPTQLPIHNQDAFARMPLMGLIG	622		
QY	1130	SDSSVLRMSDP-----	SHTTNKLYPKORVLDVSRSPETCKPQN-----	SIEND	1175
Db	623	SSPHLPNLSUPPGSGLGTFFSAIAOSSYPDAK--DKNSAFNPMDSPNNSWTSAPTVEGE	680		
QY	1176	IEEKVSLASRGQEPKTSKTKGNDFIDDSKLASADIGITLICKN-KKPLIOESDITVS	1234		
Db	681	-----	NDTMSAQRSTLKWEK-----	EELGEMATVAPVLYTNINFPNLKEEFPDWT	728
QY	1235	SSKSALHSSVPKSTNDRDATPLSRAMDFEGLGCD--	SESNSTLSENSDVTVSIDSSSE--	1290	
Db	729	RVKQ-IAKLWRKASSQERAPYVQARDNRAALINKVQMSNDMSKRRQOQDSIDPSSRID	787		
QY	1291	---EDMTVQNSNISISEQRTREQ-----	DVEVLEPLKCE-----	1322	
Db	788	SELFKDPULKRESEH-EOEWKFRQOMROKSKQQAQKIEATQKLEQVKNQEQOQOQOFGSQ	846		
QY	1323	---LVSGESTGNCEDRLPVKTEANGK-KPSQQ-----	KKLEERPYNKSDQIKLKN	1371	
Db	847	HLLVQSGSDTSSGSIQSPLTTPQNGNNSPAQSFHKLFTKQPPSTPTSTSSDDVFEVVKPQ	906		
QY	1372	T-----	DKKNENRESEKKQRTSTFQINGKDNKP-KIYLK-----	1406	
Db	907	APPPPPAPSRIPIDSLSQAQTSQPPSQVSPGSSNSRPPSPMDPYAKMVGTPRPPVPG	966		
QY	1407	-----	GCLKEISESRVYVGNVEPKVNNINKIIPENDI--KSLTVKESAIRPF	1452	
Db	967	HSFERRNSAAVENCNCTPLSSVSRPLQMN-----	ETTANRPPVRDLCSSTTNNNDPYAKPP	1022	
QY	1453	INGDVIMBDFERNSSSETKSHLLSSSDAEG-----	NYR	1485	
Db	1023	DTPRVMTDQFPKSLGLSRSPVVSQTAAGTSDHFTKPSRADVFQORIPDSYA	1082		
QY	1486	DSLETLPTKESDSTQTTTPSASCPE-----	NSVNQV-EDMEIETSEVKVKTSSPITSEE	1540	
Db	1083	RPLLT-PAPLDSGFGFKTPMQPPSPSQDPYGSQASRRLSVDPYERPALTPRPIDNFS	1141		
QY	1541	ESNLSNDFIDENGLPIKNENNVGESKRKTVITEVTTMTSTVATESKTVIKEGDKQTV	1600		
Db	1142	H-NOSNDYSP--PLTPHPAVN-----	ESFAHPSRAFSQPGTISRPTSDQPYSQ	1188	
QY	1601	VYST-----	ENCAKSTVTTTTTTTKLSTPSTGSDVSIISVKEQSKTVTTTVDLSTTT	1655	
Db	1189	PPGTPRPVVDYSQSSGTFARSNTDPYSQPGTPRPTTVDPYSQQPQTPRSTQDLFVTP	1248		
QY	1656	GGLVTSMTVSKVSTRDKVKLMKFSRPKTRSGTALPSYKFKVTKSKSIFVLPNDDL	1715		
Db	1249	---VTNQRHSDPYAHPGTPRPGISVPSYOPPATPRPRISEGETRSMTRPVLMPNQD-	1303		
QY	1716	KKLARKGIREVPYFVN--YNAKPALDIWPYSPRPTFGITWRYRLQTVKSLAGVSLMLRL	1773		
Db	1304	-----	PFLQAAQNRGPAL--PGFLVRP-----	1323	
QY	1774	LWASLRWDDMAAKVPPGGGSTRTETSETTEITTEIKRRDVGYPYGEYFECIRKICPIG	1833		
Db	1324	-----	1323		
QY	1834	VPETPKETPTQORKGLRS--SALRPKRPEPTKQGTGPVLIETWVAEELELWEIRAFARV	1891		
Db	1324	-PDCSQTPRPPGCLSDTFSRVSPSAARDPYDQSPMTPRS-----	1363		
QY	1892	EKEKAQAVEQOAKKRLEQOKP--TVIATSTTSSTSTSTISPAQKVMVAPISGSVITG	1948		
Db	1364	-QSDSFGTSQTAHDVADPRPGSGSCFACSSNPMHSQOQFGVSGV-LPGVPVTSVGTD	1421		
QY	1949	TKMVLTTKVGSPATVTFQKNKFNHQTATWVKQGSNSGVV--QVQKVLGIIPTSSGT	2005		
Db	1422	T-----	QNTVMAQADTEKLQRQKLEIILQOQOQKIKIAGQKESQD	1465	
QY	2006	S-----	QOTFT-----	SFQPR	2017
Db	1466	SPAVPHFGPQHWQFENVNQAFTRPPPPYPGNIRSPVAPPLGPRYAVFPKQGRGYPYDV	1525		
QY	2018	ATVITIRPN-----	TSGSGGTNS-----	QVITG--POLRPGMTV-----	2050
Db	1526	ASMGMRPHGFRFGPGSGHGTWPQOERFLVPPQIQGSGVSPQSRVSVDMRPLNNSQ	1585		
QY	2051	IRTEL-----	QOSTLGKAI-----	RTPVVQPGA-----	2075
Db	1586	MNNVGLPQHFSPOSTLPVOQHNLGQAVIELHRAPDGRQRLPFSAPGVSVEASSNLRH	1645		
QY	2076	-----	POQVMTQI-----	TRQQVSTVVASPNTV--	2099

Db 1646 GNFIPRPDPFGRHTDMRRPQGLPNLPVHPDLEQVPPSQEQGHSHSSMMVRTLN 1705
Qy 2100 -----SSTPGQKSLTSATSTSNIOSSA-----SQPRPQOGVK----- 2133
Db 1706 HPLGGESEAPLSTVSPSETSDNLOITTPQSDGLEKLDSDPSVKDELVDKLEGVEVK 1765
Qy 2134 -----LTMQAQTQLTQTHGGNOGLTVVIQOGQTTGQLQLIP----- 2170
Db 1766 DLDDDELENLNDTEGKVELDTLONLETNPNLDDLLR-----SGEDIIATYDPELD 1820
Qy 2171 -----QGVTVLPQPGQQLMQAAMPNGTVORFLFTPLATATTA 2208
Db 1821 MGDKSMFNEELDLPIDDKLDNQCVSEPKKEQ-----ENKTLVLSDKHSPQKK 1870
Qy 2209 STTTTIVSTTAAGTGOROSKL-----SPQOVHODKTLTPAOSSSVGPA-- 2253
Db 1871 STVTNEKTEVLSPNSKVESKETENKDNKNVDTPCSAHSNDNGEKTSLHPCDP 1930
Qy 2254 -----KAQPQTAPSPARPOTQPSAPQPEVQTOPEVOTQTVSSHVPSEAQPT---HA 2305
Db 1931 DLPEKRTNRETAGPSAN-VIQAOSTQLPAQDVINSCGITGSTPVLSSLANKEKSDNDIRP 1989
Qy 2306 QSSKPOVAQSQPOSQVQSPVRQSPQTRIRPSTPSQLSPGQ-----QSQ 2353
Db 1990 SGSPPTTLPASPSNHVSLPPF-----IAP-----PGRVLDNAMSNVTVWSR 2033
Qy 2354 VQTTTQPIPIOPHTSLQIPSQ-----GPOSQ---PQV-QSSTQTLSSGQTL---N 2398
Db 2034 VNHVFSQGVQVNPGL---IPGOSTVNHSLGTGKPATQTPQTSQTSMSGPOQLMIPQ 2090
Qy 2399 QVSVSPSPQLOIQOPQPVIAVPQLOQOVQVLSIQISOV----- 2439
Db 2091 TLAQONRERPLLEEQPLLDLDLQEROEQOQOQOMQAMIRSEPPFPNIDFADITDP 2150
Qy 2440 -----VAIOAQOS-GVPOQIKQLPLQIOQSSAVQ----- 2469
Db 2151 IMKAKMVALKINKVMAQNLMPPMVMVSRFPMPGVGTQNSQNLGPOAIPQDGS 2210
Qy 2470 THQIQN-----VVTVQAASVOEQLRVQQLRDQOQKQOQOQIEINVTNTPSKLLIK 2519
Db 2211 THQISRPNPFPNGFVNDQSQRQYEWLEQTOQLLOMOQKYLEEQIG----- 2258
Qy 2520 VEIQOVVMKHNNAVTEHLKQKSMTPADREENQRMIVCNQVMKYILDKIDKEEQAANK 2579
Db 2259 -----AHRKSKKALSAKOR-----TAKK 2276
Qy 2580 RKREESVEQKRSQONATKLSALLFKHKEQLRAELKRALDKLDLOIEVQEELKRLDKIK 2639
Db 2277 AGREFFPEDEQLKHVTEQOSQVMYQKQLEQIRK-----OQKEHAELIEDYRIK 2323
Qy 2640 KERDLMLQAQATAAAP-CPPVTPVLPAFPAPPPPPPGV-----QHTGLLSTPTL 2691
Db 2324 QQQ---QCAMAPTMPSVQVQPPPLIPGATPPTMSQPTFPMVPPQLOHQHTTVISGHTS 2380
Qy 2692 PV 2693
Db 2381 PV 2382

RESULT 9

US-09-854-133-425
; Sequence 425, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133

Query Match 2.7%; Score 402; DB 10; Length 4019;
Best Local Similarity 17.3%; Pred. No. 1.8e-11;
Matches 510; Conservative 314; Mismatches 992; Indels 1126; Gaps 108;
254 CRVCHKL---GDLCCETCSAYVHLECYKPPLEEYPEDEWQCEVCAHVPGVTCVABI 310
Db 65 CEACGKATDPGRLLLCDDCDSYHTYCLDPPLOTVPKGGWKCKWCWCRHCGATSAGLRC 124
Qy 311 QNKPKYIRHEPTGYDRSRKRYWFLNRLLIEDTENENEKKIWIYVSTKVOLAELIDCLDK 370
Db 125 EWQNNYTQCAPCA-SLSCPCVCYRNYR---EEDL-----LLOCRQC 161
Qy 371 DYWEAELCKILEEMREEIHRHMDITEDLTNKGKSNKSFLLAAANEELLESIRAKKGDIDN 430
Db 162 DRMHAVCONL-NTEEEVENVADIGFDCS-----MCRPYMPAS-----N 199
Qy 431 VKSPETEKKDKNETENDSKDAEKNEEFEDQSLKSDDKTDPDDPEQKSEEPTEVGDK 490
Db 200 VPSSDCE-----SSSLVAQIVTKVKELDPPKTYTQDGV----- 232
Qy 491 GNSVSANLGDNTNATSEETSPGSRSPVGCISEPTDSSNMAE-KKVASSELQDVPPEPN 549
Db 233 -----CLTE-----SGMTQLOSLTVTVPRKRSPK 258
Qy 550 KTCESN-TSATTTSIQPNLENSN-SSELNSQSSEAKAADDPENGERSHTPVSIQEE 607
Db 259 LXLKIINQNSVAVLQTPPDIOSEHSRDEGMDSDREGELMDCD-----GKSES-----SPERE 310
Qy 608 IVGDFTSKTSSELSPGAGKAGSGSTRITIRLNPDSKLSQKLSQQAHAHANKLF 667
Db 311 AVDDET-----KGVEGTGKVRKRKRK-----Y 333
Qy 668 KEGKEVLVYNSOGEISRLSTKKEVIMK---GNINNVFKLQEGKYRVYHNOYSTNSFALN 724
Db 334 RPIGIGFMVQRSRQRTQGGTKRSVIRKSDSGSISEQLPCRDDG----- 376
Qy 725 KHQREDHDKRRLAHKFLCTPAGEFKWNGSVHGSKVLITISTRLTITOLENNIPSSFFH 784
Db 377 ----- 376
Qy 785 PNWASHRANWIKAVQCMKPRFALALALECAVKPVVMLPIWREFLHTRHRTSIR 844
Db 377 -----WSEQLPDTLVDESVSV-T 393
Qy 845 EEKEKVKKEKKEE---EET---MOQATWVKYTFPVKHQVWKQKGEERYVTGYGWSWI 898
Db 394 ESTEKIKKRYKRKKKLEETFPAYLOEAFEGKDLDTSRQ-----SKI 436
Qy 899 SKTHVYRFVFKLPNGTNNVNRKSLGKTKNNMDENMDESCKRCSRSPKKIKIEPDEKDE 958
Db 437 SLDNLSGDAQGLLYKTNM-----TGLDPLSLD--PLLSSSSAPTK----- 475
Qy 959 VKGSDAAKADQONEMDISKITEKKDQDVKELDDSDKPKCKEPM-EVDDDMKTESHVNC 1017
Db 476 ---SGTHGPAADDPLADISEVL-NTDDDLIGIISDDLAKSVHSDIDIGVTDSSLPQPNV 531
Qy 1018 QESSQVDVNVVNSGFFHLRTSYKKKTKSSKLDGLLERRIKOFTLEEKQRLKLEKGIK 1077
Db 532 NQSS-----RPLSEQLDGLL-----SPELDKMVTDCAILG 562
Qy 1078 -IGKTSTNSKNLSSESPVITKAKEGQOSMRQEQP-----NANNQOPDELIOGCSQ 1129
Db 563 KLYKIPELGGKDVEDLFTAVLSPANTQPTPLPQPPPTQLLPIHNOQDAFSRMLMGLIG 622


```
QY 1949 TKWLTTKVGPATVTFQONKNPHOTFATWVKQGSNGV---QVQOKVLGIIPSTGT 2005
Db 1422 T-----ONTVNAQAADTEKLQRQKRLREIILQOQKQKLAGRQGSQD 1465
QY 2006 S-----QOTFT-----SEOPT 2017
Db 1466 SPAVPHGPGLOHWPENVAQAFTRPPPPYCGNIRSPVAPPLGPRYAVFPKQDQGPYPDV 1525
QY 2018 ATVTIRN-----TSGSGGTTNS-----QVITG-----PQIRPGMTV----- 2050
Db 1526 ASGMGRPHGRFPGSGSHGTMPSQERFLVPPQIQGSGVSPQLRRSVSDMPRLNNSQ 1585
QY 2051 IRTPL-----QOSTLGRALI-----RTPVWVQPGA----- 2075
Db 1586 MNNPVGVLPHQFSPQSLPVQOHNILGQAYIELHRAPDGRQLRPFSPAPPSVVEASSNLRH 1645
QY 2076 -----POQVMTQI-----IRGPVSTAVSAPNTV- 2099
Db 1646 GNFIPRPDPGPRHTDPMRPPGGLPNQLPVHPDLEQVPPSQEQGHSHVSSNMVMTLN 1705
QY 2100 -----SSTPGQKSLTSATSTNIQSSA-----SOPPRPQGGQVK----- 2133
Db 1706 HPLGGEFSEAPLSTVSPSETSDNLQITTPSDGLEKLDSDPSVKELDKVKGVEVK 1765
QY 2134 -----LTMALQLTQCGHGNGLFVVIQGGCTTGQLQILP----- 2170
Db 1766 DLDDLEDNLNLTDEGKVVYELDTLNLNETNDPNDLDDLLR-----SGEFDIAYTDPELD 1820
QY 2171 -----OGVTVLPQPGQOLMOAAMPNGTVQREFLTPLATATTA 2208
Db 1821 MGDKSMFNEELDLPIDDKLDNCVSEPKKEQ-----ENKTLVLSKHSFQKK 1870
QY 2209 STTTTIVSTTAAGTGBORQSKL-----SPQMVOHQDKTLPAPQSSSVGPA-- 2253
Db 1871 STVTNEVKTEVLSPNKSVKESKCEKTEKNDENKDNVDTPCSOASAHSDLNDGKTSLHPCDP 1930
QY 2254 -----RAQPTAOPSARPQOTQPSQAQPEVOTQPEVQTQTVVSHVSEAPQ---HA 2305
Db 1931 DLFEKRTNRETAGPSAN-VIQASTQLPAQDVINSCGITGTPVLSSLLANEKSDNSDIRP 1989
QY 2306 QSSKPOVAQSQNSVQGSPPVRVQSPQSTRIRPSTPSQLSPQ-----OSQ 2353
Db 1990 SGSPPTPLPASNSHVSLPPE-----TAP-----PGRVLDNAMNSNVTVSR 2033
QY 2354 VQTTTSOPIQPHTSLSQIPSQ-----GPOSQ--PQV-QSSTOTLSSQTL---N 2398
Db 2034 VNHVFSQGVQVNPGL---IPGQSTVNHSLGTGPKPATGTGPTQSTQSTSSSGPQQLMIPQ 2090
QY 2399 QVSVSSPSRPLQIQOPOPOVIAVPQLOQOVQLSQISOV----- 2439
Db 2091 TLAQQRNREPLLEEOPLQLLDLQERQEQOQOQOMAIQRSEFFPNIDFDAITDP 2150
QY 2440 -----VAQIQAQOS-GYPOQIKQLPQIOQSSAVQ----- 2469
Db 2151 IMKAKVALKGINKVAQNGLNPPMVMSPFPNGQVVTGTQNSQONLGPQAIPOGSI 2210
QY 2470 THQIQN-----VVTVQAASVQQLQVQRLDQOQKQKQOQIENVNTPSKLLIK 2519
Db 2211 THQISRPNPNFPGFVNSQRQYEEWLQETQOLLQMOQKYLEEQIG----- 2258
QY 2520 VEIQIQVNMKNHNAVIEHLKOKSMTPABREENQRMVCMQVNMKYILDKDKEKQAAKK 2579
Db 2259 -----AHRKSKKLSAKOR-----TAKK 2276
QY 2580 RKREESVEQKRKONATKLSALLFKHKEQLRAEILKRALDKDLQIEVQEEELKRLKIK 2639
Db 2277 AGREFFEEAEQLKHVTEQOSMVQKQLEQIRK-----OQKEHAELIEDYRIK 2323
QY 2640 KEDLMQLAQATAAAP-CPPTVPLPAPAPPSPPPPPGV-----QHTGLLSTPTL 2691
Db 2324 QOO---QCAMAPMTMPSVQOPPLIPGATPPTMPSQPTFPVWVQQLHQHQHTTVISGHTS 2380
QY 2692 PV 2693
```

Db 2381 PV 2382

RESULT 11

```
US-09-801-368-108
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
; LENGTH: 1367
; TYPE: PRM
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108
```

```
Query Match 2.7%; Score 401.5; DB 10; Length 1367;
Best Local Similarity 21.2%; Pred. No. 4.5e-12;
Matches 289; Conservative 168; Mismatches 564; Indels 339; Gaps 56;

QY 1261 DFGG---KLGCDSE-----SNSTLENSDTVSIQDSDSEDMIVQNSIESQERT 1308
Db 191 DFPFGWNIIDCNDCCGKTSTTTSSTSESSTTTSSTSESSTT 243
QY 1309 REQDVEVLEPLKCELYSGESTGNC--EDRLPVKGTENGKPKSOQKKLEERPYNKCSQI 1366
Db 244 TSSTSESSTSSSTAPATPTTJCTREKPTPTTCTCKEKPp-----PHDTPCT 296
QY 1367 KLKNTTDKKNNRESEKKQORTSTFOINGKDNKPKIYLKGECLKETSESRVSVGNVEPK 1426
Db 297 KKTTTSKTKTKTTTPVTPPSSSTTE---SSSAPVTPSSSTSSSAPVTSSTTE-- 350
QY 1427 VNNIKLIIPNDIKSLTVKESAIRPFINGDVIMEDFERNSSSETKSHLLSSDAEGNYRD 1486
Db 351 --SSSAPV---TPSSTTESAPVTS-----STTESAPVTSSTTESAP 394
QY 1487 SLETLPTSKESDS---TQTTTPSASCPSNVNOVEDMEIETSEVKKVTSSPITSE--- 1539
Db 395 VTPPSSSTTESAPVTSSTTESAPVTSSTTESAPVTSSTTESAPVTSSTTESAPVTSSTTES 453
QY 1540 -----EESN--LSNDFIDENGLPIKNKNVNGESKRKKTIVITEVTMTST-VA 1583
Db 454 SSAPVTPPSSSTTESAPVTSSTTESAPVTPPSSSTTESAPVTSSTTESAPVTSSTTESAPV 513
QY 1584 TESKTVIKVEKGDKQVWVSTENCASKSTVTTTTVTKLSTPSTGSGVDIISKEQSKTV 1643
Db 514 TPSSTTESAPVTPPSSSTTESAPVTSSTTESAPVTPSS-----STTESSTP 568
QY 1644 VTTTVVDSLTTTGGTLVTSMTVSKE--YSTRDVKLMLKFSRPPKKTGRGTALPYSYRKFTVK 1701
Db 569 VTSSTTESAPVTPPSSSTTESAPVTPPSSSTTESAPVTPPSSSTTESAPVTS 628
```

QY 1702 STKKSIFVLPNDLKKLARKGIREVPYFNNAKPALDIWPPSPRTFGITWRYRLQTV 1761
Db 629 STTESSA-----PVPTPSSS-----TTE 647
QY 1762 KSLAGVLMRLMLASLRWDDMAAKVPGGSGTRTETETEITT-----TEIKR 1811
Db 648 SSSAPVTP-----SSSTESSAPVTPSSST-TESSAPVTPSSSTESSAPVTSSTE 701
QY 1812 RDVGYPGIRFECIRKICPIGVP-----ETPKETPTPQKGLRSSAL----- 1854
Db 702 SSSAPVTPSSSTTESSAPVTPSSSTESSAPVTPSSSTESSAPVTSSTESS 761
QY 1855 -----RPKRPETPKQCPVETIETWAEELWEIRAFARVEKEKAQAQVQAKRLBQQ 1910
Db 762 APVTPSSSTTESSAPVTPSSSTTESSAPVTPSSSTTESSAPVTPSSSNIITSS 821
QY 1911 KP--TVIATSTTS-----PSSSTTISPAQKVMVAPISGVSVTGKMWLTTKVGS-- 1959
Db 822 APSSTPFSSTESSVPVTPSSSTESS-----APVSSSTTESSAPVTPSSSSN 874
QY 1960 -----PATVTFQONKNEHOTATWVKOGNSGVVQVKVLG-----LIIPS----- 2001
Db 875 ITSSAPSPFSSTTESSTEST-GTTVPSSSKYPSQTETSVSSTTETITVTKTTSTVT 933
QY 2002 -----STGTSQ--QTFTSPQRTATVIRPNTSGSGGTTSSNSQVITPQIRPCM 2048
Db 934 PSTTTITTTVCSTGNSAGETSGCSPKVTITV-PTTTTTSVTSSTTIT------T 985
QY 2049 TVIRTPLO--OSTLG--KAIIRT-PVMQPGAQVQVMTQIIRQOPSTAVS-----AP 2096
Db 986 TVCSTGNSAGETSGCSPKTTTTPVPCSTPSETASESTTSPTPVTVTVTVTVTTE 1045
QY 2097 NVVSTPGOKSLTSATSNIOSS--ASOPRPOQGV-KLTMAOLTQTOGHGNGOGLT 2153
Db 1046 YSTSKPGGE-ITTFVTKNIBPTVLTITAPSPVTVTNFTPTTITTVCTSGTNSA-- 1102
QY 2154 VVIQOGTGLQLOLIPQGVTL-----PQPGQQLMQAAMPNGTVQRFLETPLATATTAS 2209
Db 1103 -----GETTSGCS--PKTVTTVPCSTGTGEVTEA-----TTLVTTA---- 1138
QY 2210 TTTTIVTITAACTGQRQSKLSPQMVQHDKTLPAQSSSVGCPAKAQOTAPQARPQPO 2269
Db 1139 -VTTTVTTTESSTGNSAGKTGTG--YTKSVPTTYVTTLAPS----- 1178
QY 2270 TQPSAPQEVQTOPEVOTQTTVSSHVPSEAOPHTAQSCKPOVAAQSQPQSNVQGSQSPVR 2329
Db 1179 ----APVTATNAVPTITTTTSCATNAAGETTSVCSAKTIVSSASAGENTAPSAT--- 1231
QY 2330 VQSPQOTRIPST--PSQLSPGQSOVQTT-----TSOPIPIOPHTSLQIPSGQPOQPOV 2384
Db 1232 --TPVTTAI-PTTVITTESSVGTNSAGETTTGTGTSIPTTITL-IPGNSGAKNYETV 1287
QY 2385 QSTQTLSSGQTLNOVSVSSPSPRQLOIQOPQOVIATVAPLOQOQVLSQIQSOVQAOIQ 2444
Db 1288 ATAT-----NPISIKTTS-----OLATTASASSVAPV 1315
QY 2445 AQQSGVPQIKLQPLIQIQSSAVQTHQIQNV-VTVQAAS 2483
Db 1316 TSFSS-----LTGPLQASGSNAVATYSPVSISSITYQGA 1348

RESULT 12

US-10-171-311-4
; Sequence 4, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen

; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 3899
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-4
Query Match 2.6%; Score 391.5; DB 15; Length 3899;
Best Local Similarity 16.8%; Pred. No. 5.8e-11;
Matches 489; Conservative 482; Mismatches 1051; Indels 881; Gaps 114;
QY 305 DCVAET-----QKNKP--YIRHEPIGYDRSRKRYW-----FLNRLRIIE 341
Db 113 DCSSEVNGCSFVMRTGKPTNLLREEFGVDDSYSEGAQSDPTHLEMMSELAGQHEIE 172
QY 342 EDTENENKKIWIYKTKVOLAEIDLCL-----DKDYWEAEELCKIL 381
Db 173 ELNRELEEMRVYTGTEGLQQLQEFBAAIKQRDGIITQLTANLQARREKDETMREFLEIT 232
QY 382 EMRE-ETH-RHMDITEDLTNNKARGSNKSFIAAANEELIESIRAKGIDIDNVKSPETEET 439
Db 233 EQSQLQIQFQQLQASETLRNSHSTHADLQAQQLLTH-----QQ 275
QY 440 DKNETENDSKAEKNREFEFQDLEKSDDKTPDDPEQKSEETPEVDGKGNVSANLG 499
Db 276 QLEEQDHLEDYQKKKEDFTMQI--SFLQEKIKVYEMEQDKKVE----- 317
QY 500 DNTWTATSEETSPSGRSPVGCLETPDSSNMAEKVASELPODPEEPNKTCSSTNSA 559
Db 318 ----NSNKEETOEKE-----TIEELNTKIIIEEKKTLLEKDKLT 353
QY 560 TTTS-----IQNLNSNSSSELNSSQESAKAAD-----PENGERSHTPV 603
Db 354 TADKLLGELQEQIVQKNQEIKNMKLELNSKQKQESSEIKQLMGTVLEQKRNKQSO 413
QY 604 IQEEIVGDTSEKSTGEISESPGAGSGASTRIITRNPDSKLSQKSOQVAAAHQA 663
Db 414 FETDIVQRMQE-----TQKLEQLR--AEIDEMYGOQIV----- 446
QY 664 NKLFEKGEVLVNVSGEISRLSTKKEVIMKGNINNYFKLGOEGKRVYVHNOYSTNSPAL 723
Db 447 -----OMQELIRQHMAQEMKTRHKGEMENALRSYNI-----TVNEDQIKLMNVAI 495
QY 724 NKQHR-EDHDKRRH-----LAHKFCLTPAGEFKWNGSVHGSVKVLTITSLRLTIQ 773
Db 496 NELNKLQDNTNSQKEKLELGLILLEKCALQ-----RQ 529
QY 774 LENNIPSSFFHPNASHRANWIKAYOMCSKPREFALALAILCAVKKPVVMLPIWREFLGH 833
Db 530 LEDLVEELSFREQIQARQTI--AEQESKLENAHKSISTVEDLKAIEVSASESKLEL 587
QY 834 TRLHMTS-----IEREEKVKKKKQDEETMQQATWVKYTFPVKHQVWKQKE---E 886
Db 588 KHEAEVTNYKIKLEMEKKNVLDORMAESQAELELURDTQLLFSHEELSKKDELE 647
QY 887 YRVTYGGWSWISKTHVYRFVPKLPQNTNVNRYKSLEGTKNNDENMD--ESDKRKRCS 944
Db 648 HRIN-----IEKLDNGLGIHYKQIQDGLQONEMSQKTIETMQEKNLITK 691

QY	945	PKKIKIEPSEKDEVKGSAAKAGD-----QNEMDISKITEKK-----DOOVKEL---989	
Db	692	QNOILILEISKLDQOOLVNSKSEMTLQINELQKEITEILROEBKEKGTLEQVEQELQK751	
QY	990	-----LDSDSKPKKEEPMEDVDDMKTSHYVNCDESSQDVVVN1028	
Db	752	TELLEKOMKEKENDLOEKAFAEAB-NSILKDEKKTLEDMLKIHTPVSQEB-----801	
QY	1029	SEGFLHRTSYKXKTKSSKLDGLLERRIKOFTLEEKQORLEK--IKLEGGIKIGKISTNSS1086	
Db	802	-----RLIFLDSIKSKSDSWKEI-EILIEENEDLKQOCIQLENEEIEKQRNTFSFAE854	
QY	1087	KNL-----SESPVIKAKEGCOSDSMRQJ-----QSPNANNQDPEDLQGCQSQSDSV1134	
Db	855	KNFEVNYQLOBEYACLLKVKDDLESDSKNOELEYKSKLKALNEE-----899	
QY	1135	LMSDPSTHTNKLPKRDVLDDVIRSPEI-----KCPKONSIENTI1176	
Db	900	LHLQORINPTVKM--KSSVFDEDKTFVAETLEMGVEVVEKDTTELMEKLEVTKREKLE--L955	
QY	1177	EKVSDLASRGQEPKSTKTKGNDFFIDSKLASADDIGTLICKNKKPLIOESDTIVSSS1236	
Db	956	SORLSDLS---EQLKQKHGEISFLNEEVKSLQJ-----KEQVSLRCLRELEIINH1004	
QY	1237	KSALHSSVPKSTNDRDAPLPSRAMDFEGLKCDSESNSTLENSSDTVISIODSEEDMIVQ1296	
Db	1005	RA-----ENVCQSDTQVSSLL-----DGVVTWMTSRGAEGSVS1036	
QY	1297	NSNESISEQFRTREODVEVLEPLKCELVSGSTGNCE---DRLP--VKGTENGKPKPSQ1351	
Db	1037	KVNGSFGESKLTWEDKVSYFE---NMTVGESQEQOLILDLHPSVTKESLSRATQPSN1092	
QY	1352	KKLEERPVNKCSQIKLKNTTDDKKNNENRESEKRGOR-----TSTFOI-----NGKD1398	
Db	1093	DKLQ-----KELNVLKSEQNDLRL-QMEARQICLSLVYSTHVQDQVREYMENEKD1140	
QY	1399	NRPKIYKGECLKEISESRVSGVPEKVNINIKIIPENDIKSTIVKESAIRPPEING--1456	
Db	1141	-----KALCSLKEELIFA--QOEKIKELQK-IHQLEQTMKTQET-----GDEG1181	
QY	1457	---VIMEDFNERNSETKSHLLSSDAEGNVRDSLETPLPSTKESDSTQTITPSASCPE1512	
Db	1182	KPHLLIGLKQKAVSECSYFLOTLCVLGEY-----YTFALK-----1220	
QY	1513	NSVNOVEDMEIETSEVKVTSPTSEESNLSDNDFIDENGLPINKN---ENVNGESKRK1569	
Db	1221	-----EVNAEDKEN-SGDYISENEDPELDQRYREVQDQFQENMH1257	
QY	1570	TVITEVT-----TMTSTVATESKTIVIKVEKGDQKQTVVSTENCASKSTVTTTITV1619	
Db	1258	TLLNKVTEYNKLLVLOTRLSKINGQOQDGHKLFGE-----ENLPKE-----1300	
QY	1620	TKLSTPSTGGSDIISVKEQSKTVTTVTVDLSLTTGGTLVTSMTVSKEYSTRDKVKLMK1679	
Db	1301	-----ETEFLSIHQ-----MTNLEDIDVNHKSLSS1327	
QY	1680	FSRPKTRSGTALPSYKRFVTKTKKSIFVLFPNDLKKLARKGGIREVPYFNNAKPALD1739	
Db	1328	LQDLEKTKLEEQVQESLSSLOQO-----LKETEQ-----NYEAE---1364	
QY	1740	IMPYSPRPFTGITWRYRLQTVKSLAGVSLMLRLMLASLRWDDMAAKVPPGGSGSTRFETS1799	
Db	1365	-----IHCLQJ--KRLQAVS-----ESTVPP-----SLP1385	
QY	1800	ETEITTTTEIKRDRVGPGRIFREYICIRKIIICPIGVPETPKETPTPQRBKGLRSSALRPKR1859	
Db	1386	VDSWITESDAQRTMPS-----CVKKNI--DGTIEFSGFYKEETNIVKLLKEQYQE1438	
QY	1860	ETPKOTGPVIEIETWA-BEELWELWEIRAFARVEKEK-AOAVEOQAKRLEOQKPTVIAT1917	
Db	1439	QLEEEVAKVIVSNSTAFQAQOTELSRISGKENTASSQOAHAVQCOQEQHYFNEMK-----L1493	

		Matches 221; Conservative 91; Mismatches 418; Indels 254; Gaps 33;	
QY	1780 WDDMAKVPPGGGSTRTET-SETETITTEIIKRRDVGPGYRFEYCIRKIIICPIGVP--E	1836	
DB	3674 ---TTTTVTPPTPTGTGTPTTTTPTTTTTPPT-PTGQ-----TPTTTPTT	3720	
QY	1837 TPKEPTPQRKGLRSSALRP-----KRPETPKQTG-----PVIIETWAAEELELWEI	1884	
DB	3721 TTTVTPTPTPTGTGTPTTTTTPPTPTPTPTGTGTPTPTTTTPTT	3771	
QY	1885 RAFAERVEKEKAQAVEOQAKRLEQKPTVIATSTPTS---STSTISIP-----AQK	1935	
DB	3772 -----PTPTPTGTGTPTTTTPTTTTTPPTPTPTPTPTPTPTPT	3804	
QY	1936 VMVAPISGSVT-----TGKMWLTTKVGSPTATVTFQONKFNHOTFATWVKQGSNSGV	1988	
DB	3805 PTTTPTTTTTPPTPTGTGTPTTTTTPPTPTPTPTPTPTPTPTPTPTPTPT	3861	
QY	1989 VQVOQKVLGIIPSSGTSTQOFTSFQPRATVIRPNTSGSGTTSNSQVITGPIRPGM	2048	
DB	3862 -----VTPTPTGTGTPTTTTPTTTTTPPTPTPTPTPTPTPTPT	3908	
QY	2049 TVIRTPLOQSTLGLKAIIRTPVMVQGAQOQVMTQIIRGQPVSTAVSAPNTVSSTPGQKSL	2108	
DB	3909 TPTPTPTGTGTPTTTTPTTTTTPPTPTPTPTPTPTPTPTPTPTPTPTPT	3963	
QY	2109 TSATSTSNIOSSASOP-PRPQOGQVKLTMAQLTQLTQGHGGNGLTVIQQOGQTGQLQ	2167	
DB	3964 QTPPTTPTTTTTPPTPTGTGTPTTTTTPPTPTPTPTPTPTPTPTPT	4015	
QY	2168 LIPOQVTVLPFGQQLMQAAMPNGVQRFLEPLATT-----ATTASTTT	2212	
DB	4016 PTTTTTTPPTPT-----TPTGT-OTPTTPTTTTTPPTPTPTPTPTPTPT	4066	
QY	2213 TTVSTTAAGTGEORQSKLSPQMVHQDKTLPAPQSSSVCPAKAQOTAPQARPQOTQP	2272	
DB	4067 TTVTPTPTPTGTGTPT-----TPTPTTTTTPPTPTPTPTPTPTPTPT	4114	
QY	2273 QSPAQEVQTPQEVQQTTSVSHVPSEAQPTHAQSSKPKQVAAQSQPSQSNVQCSQSPRVQS	2332	
DB	4115 VPTPTPTGTGTPTTTTPTTTTTPPTPTPTPTPTPTPTPTPTPTPTPTPT	4172	
QY	2333 PSQTRIRPS---TPSQLSPGQSQVQTTTSQPIQPIQPHLSLOIPSGOQPO-----SQQV	2384	
DB	4173 PTTTPTTTTTPPTPTPTGTGTGPPHTSTAPIAELTTTNSPPPSSTPTQTSRSTSSPLT	4232	
QY	2385 QSSTQTLSSGQTLNQVSVSSPSRP 2408		
DB	4233 ESTLLSTLPPAIENTSTAPPSTP 4256		

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 09:12:14 ; Search time 23012 Seconds
(without alignments)
17537.523 Million cell updates/sec

Title: US-09-698-295-2
Perfect score: 9865
Sequence: 1 ggcaggctgagtgccgc.....acaaaaaaaaaaaaaaaaa 9865

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_da.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_hgt_hum.*
- 31: em_hgt_inv.*
- 32: em_hgt_other.*
- 33: em_hgt_mus.*
- 34: em_hgt_pln.*
- 35: em_hgt_rod.*
- 36: em_hgt_mam.*
- 37: em_hgt_vrt.*
- 38: em_sy.*
- 39: em_higo_hum.*
- 40: em_higo_mus.*
- 41: em_higo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	9067	91.9	9700	9	AB032251	AB032251 Homo sapi
2	7836	79.4	8295	9	AY282495	AY282495 Homo sapi
3	3033	30.7	3419	9	AK025776	AK025776 Homo sapi
4	2565.6	26.0	2673	9	HSU05237	U05237 Human fetal
C 5	2316	23.5	88848	9	AC107377	AC107377 Homo sapi
C 6	2314	23.5	108860	9	AC127525	AC127525 Homo sapi
C 7	2314.4	23.5	199875	9	AC134407	AC134407 Homo sapi
8	1998	20.3	4074	9	HS080468	AL833375 Homo sapi
9	1719.4	17.4	2057	6	BD157797	BD157797 Primer fo
10	1719.4	17.4	2057	9	AK022570	AK022570 Homo sapi
11	1714	17.4	2549	10	BC032289	BC032289 Mus muscu
12	1708.4	17.3	2769	10	BC046973	BC046973 Mus muscu
13	1690.2	17.1	2481	10	BC037661	BC037661 Mus muscu
14	1496.2	15.2	2214	10	BC021489	BC021489 Mus muscu
15	1454.4	14.7	23387	2	AC109873	AC109873 Rattus no
C 16	1454.4	14.7	238226	2	AC120467	AC120467 Rattus no
C 17	1454.4	14.7	280450	2	AC111698	AC111698 Rattus no
C 18	1448.4	14.7	217735	10	AL596116	AL596116 Mouse DNA
19	1394.6	14.1	1606	9	BC050566	BC050566 Homo sapi
20	1240.2	12.6	1540	6	AX552123	AX552123 Sequence
21	1190.6	12.1	1576	9	BC037422	BC037422 Homo sapi
22	1071.2	10.9	1955	10	BC003956	BC003956 Mus muscu
23	952.2	9.7	1976	9	AK027184	AK027184 Homo sapi
C 24	831	8.4	179759	2	AC132868	AC132868 Homo sapi
C 25	831	8.4	191280	2	AC087389	AC087389 Homo sapi
26	831	8.4	193519	2	AC023123	AC023123 Homo sapi
27	814.6	8.3	1032	6	AR216196	AR216196 Sequence
28	771.4	7.8	841	6	BD147323	BD147323 Primer fo
29	688.2	7.0	62138	2	AC100831	AC100831 Homo sapi
C 30	662.6	6.7	60526	2	AC130323	AC130323 Homo sapi
31	627.4	6.4	750	6	BD018361	BD018361 Novel gen
32	627.4	6.4	750	6	BD098299	BD098299 Novel gen
C 33	598.2	6.1	183093	9	AC138688	AC138688 Homo sapi
34	598.2	6.1	187685	9	AC127032	AC127032 Homo sapi
35	598.2	6.1	190285	2	AC126604	AC126604 Homo sapi
36	598.2	6.1	190844	2	AC073990	AC073990 Homo sapi
37	586.2	5.9	139585	9	AC090419	AC090419 Homo sapi
38	586.2	5.9	170030	9	AC005829	AC005829 Homo sapi
39	586.2	5.9	226696	9	AC139026	AC139026 Homo sapi
40	585.8	5.9	134016	2	AC048388	AC048388 Homo sapi
41	574	5.8	574	6	AX677732	AX677732 Sequence
42	574	5.8	2123	6	BD159315	BD159315 Primer fo
43	574	5.8	2123	9	AK022081	AK022081 Homo sapi
44	547.6	5.6	80206	2	AC138693	AC138693 Homo sapi
C 45	538.2	5.5	556	6	BD151960	BD151960 Primer fo

ALIGNMENTS

RESULT 1
AB032251
LOCUS
DEFINITION Homo sapiens BPTF mRNA for bromodomain PHD finger transcription factor, complete cds.
ACCESSION AB032251
VERSION AB032251.1 GI:6683491
KEYWORDS bromodomain PHD finger transcription factor; BPTF.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
REFERENCE Jones, M.H., Hamana, N. and Shimane, M.
AUTHORS Identification and characterization of BPTF, a novel bromodomain
TITLE

AB032251 9700 bp mRNA linear PRI 01-MAR-2000
Homo sapiens BPTF mRNA for bromodomain PHD finger transcription factor, complete cds.

Db	1115	ATGACCAGTAGAGAACAGATCAAAAGTTCTACAGTTTCTAGTCGATCAGTTCTTTACAA	1174
QY	962	CAAATATTCCTCAGAGGAATTCATGCTCTGAAGGGGTGATACAGTATGATGACCATTTGTA	1021
Db	1175	CAAATATTCCTCAGAGGAATTCATGCTCTGAAGGGGTGATACAGTATGATGACCATTTGTA	1234
QY	1022	GGTTTGTGCACAAACTTGGGGATTGCTTTGCTGTGAGACATGTTTCAGCAGATATACCAAT	1081
Db	1235	GGTTTGTGCACAAACTTGGGGATTGCTTTGCTGTGAGACATGTTTCAGCAGATATACCAAT	1294
QY	1082	TGGAATGTGTGAAGCCACCTCTTGAGAGGTGTCACAGGACGAGTGGCAGTCTGAAGTCT	1141
Db	1295	TGGAATGTGTGAAGCCACCTCTTGAGAGGTGTCACAGGACGAGTGGCAGTCTGAAGTCT	1354
QY	1142	GTGTAGCACACAAGSTGCCTGTGTGACTGACTGTGCTGAAATCCAAAAATAAAC	1201
Db	1355	GTGTAGCACACAAGSTGCCTGTGTGACTGACTGTGCTGAAATCCAAAAATAAAC	1414
QY	1202	CATATATTCGACATGAACCTATTGGATATGATAGAAAGTCGGAGGAATACTGGTCTTGA	1261
Db	1415	CATATATTCGACATGAACCTATTGGATATGATAGAAAGTCGGAGGAATACTGGTCTTGA	1474
QY	1262	ACGGAAGACTCAATATAGAGAGATACAGAAAAATGAAAATGAAAAGAAATTTGGTATT	1321
Db	1475	ACGGAAGACTCAATATAGAGAGATACAGAAAAATGAAAATGAAAAGAAATTTGGTATT	1534
QY	1322	ACAGCACAAAGSTCCAACTTGCAGAAATTAATGACTGCTAGACAAAGATTAATGGGAAG	1381
Db	1535	ACAGCACAAAGSTCCAACTTGCAGAAATTAATGACTGCTAGACAAAGATTAATGGGAAG	1594
QY	1382	CAGAACTTCGCAAAATTCAGAAAGAAATCGGTGAAGAAATCCACGACACATGGACATAA	1441
Db	1595	CAGAACTTCGCAAAATTCAGAAAGAAATCGGTGAAGAAATCCACGACACATGGACATAA	1654
QY	1442	CTGAAGACCTGACCAATAGGCTCGGGCAGTAACAAATCCTTCTCGGGCGAGCTAAATG	1501
Db	1655	CTGAAGACCTGACCAATAGGCTCGGGCAGTAACAAATCCTTCTCGGGCGAGCTAAATG	1714
QY	1502	AGAAATTTTGGAAATCCATAAGAGCCAAAGAGGAGACATTTGATAATGTTAAAGCCAG	1561
Db	1715	AGAAATTTTGGAGTCCATAAGAGCCAAAGAGGAGACATTTGATAATGTTAAAGCCAG	1774
QY	1562	AGAAACAGAAAGACAAAGATGAGACTGAGAAATGACTCTAAAGATGCTGAGAAAAACA	1621
Db	1775	AGAAACAGAAAGACAAAGATGAGACTGAGAAATGACTCTAAAGATGCTGAGAAAAACA	1834
QY	1622	GAGAAGATTTGAAGACCACTCCCTTGAAGAGACAGTACGACACAAACACCAAGATGATG	1681
Db	1835	GAGAAGATTTGAAGACCACTCCCTTGAAGAGACAGTACGACACAAACACCAAGATGATG	1894
QY	1682	ACCTTGAGCAAGGAAATCTGAGGAGCCCAACAGAAAGTTGGGATAAAGTAACTCTGTGT	1741
Db	1895	ACCTTGAGCAAGGAAATCTGAG	1917
QY	1742	CAGCAAAATCTGGCGAACACAAACAAATGCAACTTCAGAGAGACTAGTCCCTCTGAAG	1801
Db	1918		1917
QY	1802	GGAGGAGCCCTGTGGGGTGTCTCTCAGAAACCCCGATAGCAGCAACATGGCAGAGACA	1861
Db	1918		1917
QY	1862	AGGTGGCATCTGAGCTCCCGGAGTGTGCCAGAGAAACCTAACAGACATGTGAGAGCA	1921
Db	1918		1917
QY	1922	GTAACACTAGTGTACCACTACCTCCATCCAGCCTAATCTGGAACACAGTAACAGCAGCA	1981
Db	1918		1917
QY	1982	GTGAACATAATTTCTTCCAGAGTGAATCTGCTAAGGCAGCTGATGATCTCTGAAATGGAG	2041
Db			
Db	1918		1917
QY	2042	AAAGAGAAATCTCATACACCTGTCTCTATTTCAGGAAGAGATAGTAGTGTATTTCACATCGG	2101
Db	1918		1936
QY	2102	AGAAGTCCACCGGGGAGCTAAAGTGAATTCCTCGAGCTGGAAAGAGGACATCTGGGTCAA	2161
Db	1937	AGAAGTCCACCGGGGAGCTAAAGTGAATTCCTCGAGCTGGAAAGAGGACATCTGGGTCAA	1996
QY	2162	CTCGAATCATCACACGATTCGGAATCCAGATAGCAAACTTACTAGCTGAAGAGCCAGC	2221
Db	1997	CTCGAATCATCACACGATTCGGAATCCAGATAGCAAACTTACTAGCTGAAGAGCCAGC	2056
QY	2222	AGGTGCGAGCCGTGCACATGAAGCAAAATAAATTTAAGGAGGCAAGAGAGTACTGG	2281
Db	2057	AGGTGCGAGCCGTGCACATGAAGCAAAATAAATTTAAGGAGGCAAGAGAGTACTGG	2116
QY	2282	TAGTTAACTCTCAAGGAGAAATTTACGGTTGAGACCCAAAAAGGAAGTATCATGAAAG	2341
Db	2117	TAGTTAACTCTCAAGGAGAAATTTACGGTTGAGACCCAAAAAGGAAGTATCATGAAAG	2176
QY	2342	GAAATATCAACAAATTTTAAATTTGGTCAAGAGGAGTATCGGCTTACCACAATC	2401
Db	2177	GAAATATCAACAAATTTTAAATTTGGTCAAGAGGAGTATCGGCTTACCACAATC	2236
QY	2402	AATACTCCCACTTCAATTTGCTTGAATAAGCACCAGACAGAGAGACCATGATAAGA	2461
Db	2237	AATACTCCCACTTCAATTTGCTTGAATAAGCACCAGACAGAGAGACCATGATAAGA	2296
QY	2462	GAAGGCATCTTGACATAAAGTTCTGCTGACTCCAGCAGGAGTTCAAATGGAACGGTT	2521
Db	2297	GAAGGCATCTTGACATAAAGTTCTGCTGACTCCAGCAGGAGTTCAAATGGAACGGTT	2356
QY	2522	CTGTCCATGGTCCAAAGTCTTACCATATCTCTGAGACTGACTATCACCACCAATTAG	2581
Db	2357	CTGTCCATGGTCCAAAGTCTTACCATATCTCTGAGACTGACTATCACCACCAATTAG	2416
QY	2582	AAACAACTATCCCTTCTCATCTTTCATCCCACTCCGATCACATAGGGCAAAATGGA	2641
Db	2417	AAACAACTATCCCTTCTCATCTTTCATCCCACTCCGATCACATAGGGCAAAATGGA	2476
QY	2642	TCAGGCAGTTCAGATGTAGCAAAACCCAGAGAAATTTGCATTTGGCTTTAGCCATTTGG	2701
Db	2477	TCAGGCAGTTCAGATGTAGCAAAACCCAGAGAAATTTGCATTTGGCTTTAGCCATTTGG	2536
QY	2702	AGTGTCACTTAAACCACTTGTGATGCTACCAATATGGCGAGAAATTTTAGGACATACCA	2761
Db	2537	AGTGTCACTTAAACCACTTGTGATGCTACCAATATGGCGAGAAATTTTAGGACATACCA	2596
QY	2762	GGTTACACCGGATGACATCAATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2821
Db	2597	GGTTACACCGGATGACATCAATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2656
QY	2822	AGAAACAG	2881
Db	2657	AGAAACAG	2716
QY	2882	TTAAGCATCAGGTTTGGAAACAAAAAGGTGAAGAGTACAGAGTACAGAGATATGGTGGTT	2941
Db	2717	TTAAGCATCAGGTTTGGAAACAAAAAGGTGAAGAGTACAGAGTACAGAGATATGGTGGTT	2776
QY	2942	GGAGCTGGATTAGTAAACTCATGTTTATAGGTTTCTTCTAAATTTGCCAGCAATACTA	3001
Db	2777	GGAGCTGGATTAGTAAACTCATGTTTATAGGTTTCTTCTAAATTTGCCAGCAATACTA	2836
QY	3002	ATGTGAATTACAGAAAGTCTGTAGAGGAACCAAAAAATATATGGAATGAAATATGGATG	3061
Db	2837	ATGTGAATTACAGAAAGTCTGTAGAGGAACCAAAAAATATATGGAATGAAATATGGATG	2896
QY	3062	AGTCACATAAAGAAAAATGTTACAGAGTCCAAAAAATAAATAATAGAGCCCTGATCTCG	3121
Db	2897	AGTCACATAAAGAAAAATGTTACAGAGTCCAAAAAATAAATAATAGAGCCCTGATCTCG	2956

Qy 3122 AAAAGATGAGGTAAAGGTTTCAGATGCTGCCAAAAGGACGACGACCAAAATGAAATGGATA 3181
 Db 2957 AAAAGATGAGGTAAAGGTTTCAGATGCTGCCAAAAGGACGACGACCAAAATGAAATGGATA 3016
 Qy 3182 TCTCAAGATTACTGAGAAGGACCAAGATGTAAGGAGCTTTAGATTCTGCAGATG 3241
 Db 3017 TCTCAAGATTACTGAGAAGGACCAAGATGTAAGGAGCTTTAGATTCTGCAGATG 3076
 Qy 3242 ATAAACCTGCAAGGAAGAACCAATGGAAGTAGACGATGACATGAAACAGAGTCCACATG 3301
 Db 3077 ATAAACCTGCAAGGAAGAACCAATGGAAGTAGACGATGACATGAAACAGAGTCCACATG 3136
 Qy 3302 TAAATTGTCAGGAGGTTCTCAAGTAGATGTGTCTCAATGTACTGAGGGTTTTCATCTAA 3361
 Db 3137 TAAATTGTCAGGAGGTTCTCAAGTAGATGTGTCTCAATGTACTGAGGGTTTTCATCTAA 3196
 Qy 3362 GGACTAGTTACAAAAGAAAACAAAATCATCCAACTAGATGGACTTCTTTGAAAGGAGAA 3421
 Db 3197 GGACTAGTTACAAAAGAAAACAAAATCATCCAACTAGATGGACTTCTTTGAAAGGAGAA 3256
 Qy 3422 TTAACAGTTTACACTGGAAGAAAACAGCGACTCGAAAAAATCAAGTTGGAGGGTGGAA 3481
 Db 3257 TTAACAGTTTACACTGGAAGAAAACAGCGACTCGAAAAAATCAAGTTGGAGGGTGGAA 3316
 Qy 3482 TTAAGGTTATAGGAAGACTTCTACAAATCTTCAAAAATCTCTCTGAATCACCAGTAA 3541
 Db 3317 TTAAGGTTATAGGAAGACTTCTACAAATCTTCAAAAATCTCTCTGAATCACCAGTAA 3376
 Qy 3542 TAACGAAAAGAAAAGGGTGTCCAGAGTGAATGAGACAAAGACAGAGCCCAATG 3601
 Db 3377 TAACGAAAAGAAAAGGGTGTCCAGAGTGAATGAGACAAAGACAGAGCCCAATG 3436
 Qy 3602 CAATATATGATCAACCTGAGACTGTGATCAGGATGTTCAAGATGTTCAAGTTCCTCAGTTC 3661
 Db 3437 CAATATATGATCAACCTGAGACTGTGATCAGGATGTTCAAGATGTTCAAGTTCCTCAGTTC 3496
 Qy 3662 TTAGAATGAGTGATCCTAGTGCATACCAACAACTTTATCCAAAAGATGAGTGTAG 3721
 Db 3497 TTAGAATGAGTGATCCTAGTGCATACCAACAACTTTATCCAAAAGATGAGTGTAG 3556
 Qy 3722 ATGATGTCCTATTCGAGGCCAGAACAAAATGTCGAAAACAAAATTCATTAAGATG 3781
 Db 3557 ATGATGTCCTATTCGAGGCCAGAACAAAATGTCGAAAACAAAATTCATTAAGATG 3616
 Qy 3782 ACATAGAAAAGTCTCTGACCTGCGAGTGCAGGCGAGGACCCCACTAAGAGTAAAA 3841
 Db 3617 ACATAGAAAAGTCTCTGACCTGCGAGTGCAGGCGAGGACCCCACTAAGAGTAAAA 3676
 Qy 3842 CCAAGGAAATGATTTTTTCATGATGACTCTAACTAGCCAGTGCAGATGATTTGGTA 3901
 Db 3677 CCAAGGAAATGATTTTTTCATGATGACTCTAACTAGCCAGTGCAGATGATTTGGTA 3736
 Qy 3902 CTTTGATCTGTAAGAACAAAACCGCTCATACAGAGGAAGTGACACCACTGTTTCTT 3961
 Db 3737 CTTTGATCTGTAAGAACAAAACCGCTCATACAGAGGAAGTGACACCACTGTTTCTT 3796
 Qy 3962 CTTTCAAGAGTCTTTACATTCATCAGTGCCTTAAAGTACCAATGACAGAGATGCCAC 4021
 Db 3797 CTTTCAAGAGTCTTTACATTCATCAGTGCCTTAAAGTACCAATGACAGAGATGCCAC 3856
 Qy 4022 CTCCTGTCGAAGCAATGGACTTTGAAGGAAAACCTGGGATGAGTCTGTAATTAATAGCA 4081
 Db 3857 CTCCTGTCGAAGCAATGGACTTTGAAGGAAAACCTGGGATGAGTCTGTAATTAATAGCA 3916
 Qy 4082 CTTTGAATAATGTTCTGTATCCGCTCTATTTCAGATAGCAGTGAAGAAGATGATTG 4141
 Db 3917 CTTTGAATAATGTTCTGTATCCGCTCTATTTCAGATAGCAGTGAAGAAGATGATTG 3976
 Qy 4142 TTCAGAAATAGCAATGAAGCAATTTCTGAACAGTTCAGAACTCGAGAACAGATGTTGAG 4201
 Db 3977 TTCAGAAATAGCAATGAAGCAATTTCTGAACAGTTCAGAACTCGAGAACAGATGTTGAG 4036

Qy 4202 TCTTGGAGCCGTTAAAGTGTGAGTTGGTTCTGTGAGTCCACTGGAAACTGTGAGGACA 4261
 Db 4037 TCTTGGAGCCGTTAAAGTGTGAGTTGGTTCTGTGAGTCCACTGGAAACTGTGAGGACA 4096
 Qy 4262 GCTGCCGGTCAAGGGGACTGAGCAAAATGGTAAAAAACCAAGTCAGCAGAGAAGAAATTAG 4321
 Db 4097 GCTGCCGGTCAAGGGGACTGAGCAAAATGGTAAAAAACCAAGTCAGCAGAGAAGAAATTAG 4156
 Qy 4322 AGGAGAGACCACTTAATAATCTAGTGAATCAAAATAAAGCTAAAAATACCACTGCAAAA 4381
 Db 4157 AGGAGAGACCACTTAATAATCTAGTGAATCAAAATAAAGCTAAAAATACCACTGCAAAA 4216
 Qy 4382 AGAATAATGAAAATCGAGAGTCTGAAAAGAGACAGAGACAGATACATTTCAAAATA 4441
 Db 4217 AGAATAATGAAAATCGAGAGTCTGAAAAGAGAGACAGAGACAGATACATTTCAAAATA 4276
 Qy 4442 ATGGAAGAATATAAACCCAAAATAATTTGAAAAGGTGAATGCTTGAAGAATAATTTCTG 4501
 Db 4277 ATGGAAGAATATAAACCCAAAATAATTTGAAAAGGTGAATGCTTGAAGAATAATTTCTG 4336
 Qy 4502 AGAGTAGAGTGTAGTAAAGTGTAAATGTTGAACCAAGGTTAAATAATAATAAATAATCC 4561
 Db 4337 AGAGTAGAGTGTAGTAAAGTGTAAATGTTGAACCAAGGTTAAATAATAATAAATAATCC 4396
 Qy 4562 CTGAGAAATGATTAATAATCATTTGACTGTTAAAGAAATCTGCTATAAGGCCATTCATTAATG 4621
 Db 4397 CTGAGAAATGATTAATAATCATTTGACTGTTAAAGAAATCTGCTATAAGGCCATTCATTAATG 4456
 Qy 4622 GTGATGTCATCATGGAAGATTTTAATGAAAGAACAGCTCCGAAACAAAATGCAATTTGC 4681
 Db 4457 GTGATGTCATCATGGAAGATTTTAATGAAAGAACAGCTCCGAAACAAAATGCAATTTGC 4516
 Qy 4682 TGAGTTCTTCAGATGCTGAAGGTAATACCGAGATAGCTTTGAGACCTTGCCATCAACCA 4741
 Db 4517 TGAGTTCTTCAGATGCTGAAGGTAATACCGAGATAGCTTTGAGACCTTGCCATCAACCA 4576
 Qy 4742 AAGAGTCTGACGTACACAGACGACCAACCCCTCAGCATCTTCTCCAGAAAGCAATTCAG 4801
 Db 4577 AAGAGTCTGACGTACACAGACGACCAACCCCTCAGCATCTTCTCCAGAAAGCAATTCAG 4636
 Qy 4802 TTAATCAGGTAGAGATATGGAATAGAAACCTCAGAAAGTTAAGAAAGTTACTTTCATCAC 4861
 Db 4637 TTAATCAGGTAGAGATATGGAATAGAAACCTCAGAAAGTTAAGAAAGTTACTTTCATCAC 4696
 Qy 4862 CTATTACTTCTGAGAGGAATCTTAATCTCAGTAATGACTTTTATGATGAAAATGGTCTGC 4921
 Db 4697 CTATTACTTCTGAGAGGAATCTTAATCTCAGTAATGACTTTTATGATGAAAATGGTCTGC 4756
 Qy 4922 CCATCAACAAAATGAAAATGTCAATGGAGAAATCTTAAAGAAAACCGTGCATCACAGAAG 4981
 Db 4757 CCATCAACAAAATGAAAATGTCAATGGAGAAATCTTAAAGAAAACCGTGCATCACAGAAG 4816
 Qy 4982 TCACACGATGACCTCCACAGTGGCCACAGAAATCAAAAACCTGATCAAGTAGAAAAAG 5041
 Db 4817 TCACACGATGACCTCCACAGTGGCCACAGAAATCAAAAACCTGATCAAGTAGAAAAAG 4876
 Qy 5042 GCGATAAGCAAACTGTGGTTCTTCCACAGAAAATCTGCAAAAATCCACTGTGCACAAACCA 5101
 Db 4877 GCGATAAGCAAACTGTGGTTCTTCCACAGAAAATCTGCAAAAATCCACTGTGCACAAACCA 4936
 Qy 5102 CCAGTACAAACAGTGCACCAAGCTTTCCACACCCCTCCACAGGCGGAGTGTGGACATCATCT 5161
 Db 4937 CCAGTACAAACAGTGCACCAAGCTTTCCACACCCCTCCACAGGCGGAGTGTGGACATCATCT 4996
 Qy 5162 CTGTAAGGAGCAGAGCAAAACCGTGTCCACAGCAGAGTGCACAGACTCCCTGACCACCA 5221
 Db 4997 CTGTAAGGAGCAGAGCAAAACCGTGTCCACAGCAGAGTGCACAGACTCCCTGACCACCA 5056
 Qy 5222 CGGAGGACACTGTTTACATCTATGACTGTGAGCAAAAGAGTATTCCACAGCAGACAAAG 5281
 Db 5057 CGGAGGACACTGTTTACATCTATGACTGTGAGCAAAAGAGTATTCCACAGCAGACAAAG 5116
 Qy 5282 TGAACACTGATGAAATTTTCAAGACCAAAAGAGACTGCTTTCAGGTACAGCTTCGCCATCCT 5341

|||||
5117 TGAACAGTGAATAATTTCAAGACCAAGAAAGACTCGTTTCAGGTACAGCTCTGCCATCCT 5176
QY
5342 ATAGAAAAATTTGTTACCAAGAGACCAACAAAGAGCAATTTTGTGTTTGGCCTAATGATGACT 5401
Db
5177 ATAGAAAAATTTGTTACCAAGAGACCAACAAAGAGCAATTTTGTGTTTGGCCTAATGATGACT 5236
QY
5402 TAAAAAGTTGGCCCCGAAAAGGAGGAATCCGAGAGGTCCCTTATTTAATTAACAATGCAA 5461
Db
5237 TAAAAAGTTGGCCCCGAAAAGGAGGAATCCGAGAGGTCCCTTATTTAATTAACAATGCAA 5296
QY
5462 AACCTGCTTTGGATATATGGCCATATCCCTTCCTAGACCGACCTTTGGCATCACTTGA 5521
Db
5297 AACCTGCTTTGGATATATGGCCATATCCCTTCCTAGACCGACCTTTGGCATCACTTGA 5356
QY
5522 GGTATAGACTTCAGACAGTAAAGTCCCTTAGCTGGAGTGAGCCTGATGTTACGGTTACTGT 5581
Db
5357 GGTATAGACTTCAGACAGTAAAGTCCCTTAGCTGGAGTGAGCCTGATGTTACGGTTACTGT 5416
QY
5582 GGGCAAGTTTGAGATGGGATGATATGGCGGCCAAGGTTCCCTCAGGAGGAGGATPAC 5641
Db
5417 GGGCAAGTTTGAGATGGGATGATATGGCGGCCAAGGTTCCCTCAGGAGGAGGATPAC 5476
QY
5642 GGCAGAAACATCCGAAACTGAAATCACAACACAGAAATATTAAGAGGAGAGATGTTG 5701
Db
5477 GGCAGAAACATCCGAAACTGAAATCACAACACAGAAATATTAAGAGGAGAGATGTTG 5536
QY
5702 GTCCCTTATGGCATTCGATTTGAATATTGTATCAGGAAATCATTTCTCCATTGGAGTTC 5761
Db
5537 GTCCCTTATGGCATTCGATTTGAATATTGTATCAGGAAATCATTTCTCCATTGGAGTTC 5596
QY
5762 CAGAAACACCAAGAAAGCCGTACACCTCAGAGGAAAGCCCTTCGATCAAGTGCACCTG 5821
Db
5597 CAGAAACACCAAGAAAGCCGTACACCTCAGAGGAAAGCCCTTCGATCAAGTGCACCTG 5656
QY
5822 GGCCAAAGAGACCAAGAAAGCCCAAGCAACTGGCCCTGTTATTTAAGAACTGGGTAG 5881
Db
5657 GGCCAAAGAGACCAAGAAAGCCCAAGCAACTGGCCCTGTTATTTAAGAACTGGGTAG 5716
QY
5882 CAGAAGAAACTGGAATTTGTGGAGATCAGGGCATTTGCTGAGAGAGTGGAGAAGAAA 5941
Db
5717 CAGAAGAAACTGGAATTTGTGGAGATCAGGGCATTTGCTGAGAGAGTGGAGAAGAAA 5776
QY
5942 AGGCACAAGCAGTTGAGCAACAGGCTAAGAAACGACTGGAGCAGAGAGCCGACAGTGA 6001
Db
5777 AGGCACAAGCAGTTGAGCAACAGGCTAAGAAACGACTGGAGCAGAGAGCCGACAGTGA 5836
QY
6002 TTGCAACTTCCACTACTTCCCAACAAGCAGTACACACGACCACTCTCTCCAGACAGA 6061
Db
5837 TTGCAACTTCCACTACTTCCCAACAAGCAGTACACACGACCACTCTCTCCAGACAGA 5896
QY
6062 AAGTTATGTTGGCCCCCATTAAGTGGCTCAGTTTACAACTGGAACCAAAATGGTACTAACTA 6121
Db
5897 AAGTTATGTTGGCCCCCATTAAGTGGCTCAGTTTACAACTGGAACCAAAATGGTACTAACTA 5956
QY
6122 CTAAGTTGGATCTCCAGCTACAGTAACTATCCCAACAAAACAGAACTTTCATCAAACT 6181
Db
5957 CTAAGTTGGATCTCCAGCTACAGTAACTATCCCAACAAAACAGAACTTTCATCAAACT 6016
QY
6182 TTGCTACATGGGTTAAGCAAGGCCAGTCAAATTCAGGCGCTGTTCAAGTACAGCAGAAAG 6241
Db
6017 TTGCTACATGGGTTAAGCAAGGCCAGTCAAATTCAGGCGCTGTTCAAGTACAGCAGAAAG 6076
QY
6242 TCCTGGGTATCATTCATCAAGTACAGGTACCAGTCAGCAAACTTTACTTTCATTCAGC 6301
Db
6077 TCCTGGGTATCATTCATCAAGTACAGGTACCAGTCAGCAAACTTTACTTTCATTCAGC 6136
QY
6302 CCAGGACAGCAACAGTACAAATTTAGGCCCAATACCTCAGGCTCTGGAGAAACCAAGCA 6361
Db
6137 CCAGGACAGCAACAGTACAAATTTAGGCCCAATACCTCAGGCTCTGGAGAAACCAAGCA 6196
QY
6362 ATTCAAGTAAATCAGAGGCCCTCAGATTCGCCCTGGGTATGACCGTGTATTAGAACACCA 6421
Db
|||||

Db
6197 ATTCAACAAGTAATACAGAGGCCCTCAGATTGCGCCCTGGTATGACCGGTATTAGAACACCAC 6356
QY
6422 TCCAAACAGTCAACACTAGGAAAGGCAATTAATTCGAACACACTGTGTATGGGTACAGCCAGGTG 6481
Db
6257 TCCAAACAGTCAACACTAGGAAAGGCAATTAATTCGAACACACTGTGTATGGGTACAGCCAGGTG 6316
QY
6482 CTCCTCAGCAAGTGATGACTCAAAATCATCAGGGGGGAGCGCTGTCTCCACTGCAAGTCTCCG 6541
Db
6317 CTCCTCAGCAAGTGATGACTCAAAATCATCAGGGGGGAGCGCTGTCTCCACTGCAAGTCTCCG 6376
QY
6542 CCCTTAACACGGTTTCTTCAACACCTTGGCAGAAAAGCTTAACCTTCAGCAACCTCCACTTT 6601
Db
6377 CCCTTAACACGGTTTCTTCAACACCTTGGCAGAAAAGCTTAACCTTCAGCAACCTCCACTTT 6436
QY
6602 CAAATATACAGTCTTCAGGCTCACAACCCCTCGCCCTCAACAAAGGACAAAGTGAAGTCA 6661
Db
6437 CAAATATACAGTCTTCAGGCTCACAACCCCTCGCCCTCAACAAAGGACAAAGTGAAGTCA 6496
QY
6662 CCATGCTCAACTTACTCAGTTAAACACAGGGCCACGGTGGCAATCAAGGTTTGACAGTAG 6721
Db
6497 CCATGCTCAACTTACTCAGTTAAACACAGGGCCACGGTGGCAATCAAGGTTTGACAGTAG 6556
QY
6722 TAATTCAGGACAAAGTCAAACTACTGGACAGTTGCAGTTGTATACCTCAAGGGGTGACTG 6781
Db
6557 TAATTCAGGACAAAGTCAAACTACTGGACAGTTGCAGTTGTATACCTCAAGGGGTGACTG 6616
QY
6782 TACTCCAGGCCCAGGCCAGCAGCTAATGCAAGCTCAATGCCAAATGGTACTGTTTCAGC 6841
Db
6617 TACTCCAGGCCCAGGCCAGCAGCTAATGCAAGCTGCAATGCCAAATGGTACTGTTTCAGC 6676
QY
6842 GATTCCTCTTTACCCCATTTGGCAACACAGCCACCAGCCAGCAGCAGCACCACCACTG 6901
Db
6677 GATTCCTCTTTACCCCATTTGGCAACACAGCCACCAGCCAGCAGCAGCACCACCACTG 6736
QY
6902 TTTCCACGACAGCAGCAGGTACAGGTGAACAAGGCAGAGTAACCTGCACCCACAGATGC 6961
Db
6737 TTTCCACGACAGCAGCAGGTACAGGTGAACAAGGCAGAGTAACCTGCACCCACAGATGC 6796
QY
6962 AGGTACATCAAGACAAAACCCCTGCCACAGCTCAGTCATCAAGTGTGGGTCCAGCAAAAAG 7021
Db
6797 AGGTACATCAAGACAAAACCCCTGCCACAGCTCAGTCATCAAGTGTGGGTCCAGCAAAAAG 6856
QY
7022 CCAGCCACAGACTGCTCAGCCTTTCAGCTGGGCCCCAGCCCAAAACCCAGCCCACTGCC 7081
Db
6857 CCAGCCACAGACTGCTCAGCCTTTCAGCTGGGCCCCAGCCCAAAACCCAGCCCACTGCC 6916
QY
7082 CAGCTCAGCTTGAAGTTTCAGACTCAGCCTGAAGTTTCAGACCCCAAAACAACTGTTTCATCCC 7141
Db
6917 CAGCTCAGCTTGAAGTTTCAGACTCAGCCTGAAGTTTCAGACCCCAAAACAACTGTTTCATCCC 6976
QY
7142 ATGTCCCTTCTGAAGCACAACCCACCCAGCAGCTCATCCAAAGCCCAAGTTGCGAGC 7201
Db
6977 ATGTCCCTTCTGAAGCACAACCCACCCAGCAGCTCATCCAAAGCCCAAGTTGCGAGC 7036
QY
7202 AGTCTCAGCCTCAAGTAAATGTTCCAAAGGACAGTCTCCTGTTGTTGTTCCAAAGTCCATC 7261
Db
7037 AGTCTCAGCCTCAAGTAAATGTTCCAAAGGACAGTCTCCTGTTGTTGTTCCAAAGTCCATC 7096
QY
7262 AGACTCGAATACGTTCCATCAACTCCATCCCAACTGTTCTCTGTCGACAACTCCAGGTTTC 7321
Db
7097 AGACTCGAATACGTTCCATCAACTCCATCCCAACTGTTCTCTGTCGACAACTCCAGGTTTC 7156
QY
7322 AGACTCAACCTCACAACCGATTTCCAATTTCAACACATACATCTCTTCAGATACCTTCCC 7381
Db
7157 AGACTCAACCTCACAACCGATTTCCAATTTCAACACATACATCTCTTCAGATACCTTCCC 7216
QY
7382 AAGGCAGCCACAGTCACAAACCCAGGTACAGTCTTCAACTCAAACTCTTTTCATCAGGAC 7441
Db
7217 AAGGCAGCCACAGTCACAAACCCAGGTACAGTCTTCAACTCAAACTCTTTTCATCAGGAC 7276
QY
7442 AAACCTTTAATCAAGTTAGTGTTCATCCCATCCCTCCCTCAGCTACAAATACAGCAGC 7501
Db
7277 AAACCTTTAATCAAGTTAGTGTTCATCCCATCCCTCCCTCAGCTACAAATACAGCAGC 7336

QY	7502	CACAGCCCCAAGTCATTGCTGCTCAGCTGCAACAAACAAAGTCCAGGTTCTCTCTCAGA	7561	8581	CAGTCCCTAGTCAACAGAGGATGCCATGACAGTCTCAGCCACCTAACAGAGAAAGATTAT	8640
DB	7337	CACAGCCCCAAGTCATTGCTGCTCAGCTGCAACAAACAAAGTCCAGGTTCTCTCTCAGA	7396	8416	CAGTCCCTAGTCAACAGAGGATGCCATGACAGTCTCAGCCACCTAACAGAGAAAGATTAT	8475
QY	7562	TCCAGTCACAGGTTGCTGCTCAGATACAGGCTCAGCAAAAGTGGTGGCCCGCAACATCA	7621	8641	GAGGCGTTGAAGAGGTTGCTCGTTCCTTACAGGCCCATTAAGATGGCTGGCCCTTCCCTT	8700
DB	7397	TCCAGTCACAGGTTGCTGCTCAGATACAGGCTCAGCAAAAGTGGTGGCCCGCAACATCA	7456	8476	GAGGCGTTGAAGAGGTTGCTCGTTCCTTACAGGCCCATTAAGATGGCTGGCCCTTCCCTT	8535
QY	7622	AACCTCCAGTTACCTTATCCAAATTCAGCAAGCAGTGTGTGCAGACTCCAGCATTCAGA	7681	8701	GAACAGTAGACCCCTAATGATGCACACAGATTAATATGGTGGTTATTAAGGAACCTATGGAC	8760
DB	7457	AACCTCCAGTTACCTTATCCAAATTCAGCAAGCAGTGTGTGCAGACTCCAGCATTCAGA	7516	8536	GAACAGTAGACCCCTAATGATGCACACAGATTAATATGGTGGTTATTAAGGAACCTATGGAC	8595
QY	7692	ATGTGGTTACAGTCAGCGCCAGTGTGCAAGAGCAGTTGCAAGGGTTTCACCAACTCA	7741	8761	CTTGCCACCATGGAAGAAAGACTACAAAGACCATATATGAAAAGCTGACGGAATTTGTG	8820
DB	7517	ATGTGGTTACAGTCAGCGCCAGTGTGCAAGAGCAGTTGCAAGGGTTTCACCAACTCA	7576	8596	CTTGCCACCATGGAAGAAAGACTACAAAGACCATATATGAAAAGCTGACGGAATTTGTG	8655
QY	7742	GGGATCAGCAGCAAAAGAAAGACAGCAACAGATAGAAATTAACCTGGAACACACCCCTCC	7800	8821	GCAGATATCACCACCAATTTTTCATTAACCTGCTTACTACAAATCCAAAGTGACTCCCACTT	8880
DB	7577	GGGATCAGCAGCAAAAGAAAGACAGCAACAGATAGAAATTAACCTGGAACACACCCCTCC	7636	8715	GCAGATATGACCAAAATTTTTCATTAACCTGCTTACTACAAATCCAAAGTGACTCCCACTT	8715
QY	7801	AAGCTTCTTAATCAAAGTTGAATCATTCAGAAACAGTGGTGTGATGAAGCATATGCTGTA	7860	8881	TACCACTGTGCAGAAAGTTCTCGAATCATTTCTGTACAGAAATTTGAAGGCTTCAAGCT	8940
DB	7637	AAGCTTCTTAATCAAAG-TGAATCATTCAGAAACAGTGGTGTGATGAAGCATATGCTGTA	7695	8716	TACCACTGTGCAGAAAGTTCTCGAATCATTTCTGTACAGAAATTTGAAGGCTTCAAGCT	8775
QY	7861	ATAGAACATTTAAACAGAAAGAGATGACTCCAGCTGAAAGAGAGAGAAATCAAAGA	7920	8941	AGCAGTCTCATTAACACAAACTGCAGTCTACAGTCTCTTAAAGTTACAGCTGTTAACT	9000
DB	7696	ATAGAACATTTAAACAGAAAGAGATGACTCCAGCTGAAAGAGAGAGAAATCAAAGA	7755	8776	AGCAGTCTCATTAACACAAACTGCAGTCTACAGTCTCTTAAAGTTACAGCTGTTAACT	8835
QY	7921	ATGATTTGCTGTAAACAGGTTGATGAAGTATATTTTGGTAATAGATAGATAAAGAGAAAA	7980	9001	AACATAAAACACACAGCAAGATCTGGTGTCTGAACATTTTAAATTAAGGAGCCAGATGT	9060
DB	7756	ATGATTTGCTGTAAACAGGTTGATGAAGTATATTTTGGTAATAGATAGATAAAGAGAAAA	7815	8836	AACATAAAACACACAGCAAGATCTGGTGTCTGAACATTTTAAATTAAGGAGCCAGATGT	8895
QY	7981	CAGGACAGCAAAACGGAAGCGTCAAGAGAGTGTGAGAGAGAAAGCTGAGCAAGCAAT	8040	9061	TTTTAGTCAGGCTATCCTGCACAGACTTGACCTTAAACTTCGTTTTTATTTGTCATAAACAG	9120
DB	7816	CAGGACAGCAAAACGGAAGCGTCAAGAGAGTGTGAGAGAGAAAGCTGAGCAAGCAAT	7875	8896	TTTTAGTCAGGCTATCCTGCACAGACTTGACCTTAAACTTCGTTTTTATTTGTCATAAACAG	8955
QY	8041	GCACCTAAGCTGTACGCTCTGCTCTTCAAGCACAAAGAGAGCTCAGAGCCGAGATCCTG	8100	9121	TCCAAATATATCTTGGCCCAATTTTGTCCAAACGGACAAAGAAAAAGCAAAAGTCAACGACA	9180
DB	7876	GCACCTAAGCTGTACGCTCTGCTCTTCAAGCACAAAGAGAGCTCAGAGCCGAGATCCTG	7935	8956	TCCAAATATATCTTGGCCCAATTTTGTCCAAACGGACAAAGAAAAAGCAAAAGTCAACGACA	9015
QY	8101	AGAAGAGAGACTCCTGGACAAAGATCTGCNAATTAAGTGCAGAGAGAGCTGAAGAGA	8160	9181	CCATTAATCTTGTCAAGATCAGATGGTTTTTACTATTGTGGCAGAGAGGAGAAAACTTTGTT	9240
DB	7936	AGAAGAGAGACTCCTGGACAAAGATCTGCNAATTAAGTGCAGAGAGAGCTGAAGAGA	7995	9016	CCATTAATCTTGTCAAGATCAGATGGTTTTTACTATTGTGGCAGAGAGGAGAAAACTTTGTT	9075
QY	8161	GACCTGAAATTAAGAAAGAAAAAGACCTGTAGTGGCTCAGGCCACACAGCAGTAGCT	8220	9241	TATTTGAAAAAAGAAAAAGAAAAAGCAAGAAAAAGATACTATGGGGTCAAGTGAACCT	9300
DB	7996	GACCTGAAATTAAGAAAGAAAAAGACCTGTAGTGGCTCAGGCCACACAGCAGTAGCT	8055	9076	TATTTGAAAAAAGAAAAAGAAAAAGCAAGAAAAAGATACTATGGGGTCAAGTGAACCT	9135
QY	8221	GCACCTGCCCCCAGTGACACACAGTTCCTTCAGGCCCTCCAGCCCTCCACCTTCACCT	8280	9301	CCATGGAAATGCCACGCTCTGCTCTCAGTGAAGAGCTGGTTTAGAGTCTCAGAGAAAC	9360
DB	8056	GCACCTGCCCCCAGTGACACACAGTTCCTTCAGGCCCTCCAGCCCTCCACCTTCACCT	8115	9136	CCATGGAAATGCCACGCTCTGCTCTCAGTGAAGAGCTGGTTTAGAGTCTCAGAGAAAC	9195
QY	8281	CCCCCTCCACCTGGTGTGCAACACACAGCGCTTCTGCCACGCCACCTTACCTTCTGCT	8340	9361	TTTTGACTGTATTTATTTATTTGTTGCAAAAAGACGCTTTTTTATTTGCTGCCCTCAATTG	9420
DB	8116	CCCCCTCCACCTGGTGTGCAACACACAGCGCTTCTGCCACGCCACCTTACCTTCTGCT	8175	9196	TTTTGACTGTATTTATTTATTTGTTGCAAAAAGACGCTTTTTTATTTGCTGCCCTCAATTG	9255
QY	8341	TCCAGAGAGAGGAGCGGGAAGAGGAAAAAGACTCCAGCTCAAGTCCAAAGAAAAAGAAA	8400	9421	TCAGCTAAGTATTTTCTTATAAAATCCAGCCCGGTTACATATAATCATCTGTATCTT	9480
DB	8176	TCCAGAGAGAGGAGCGGGAAGAGGAAAAAGACTCCAGCTCAAGTCCAAAGAAAAAGAAA	8235	9256	TCAGCTAAGTATTTTCTTATAAAATCCAGCCCGGTTACATATAATCATCTGTATCTT	9315
QY	8401	ATGATCTCTACTACCTCAAGGAACTAAGAGGACACAAAGCTTTACTGTATCTGTAAA	8460	9481	ATCATGATTCCTGTAGGTAAAGTACAGAGCAGCTCTAGATGCTTTCTTCTCTATGAA	9540
DB	8236	ATGATCTCTACTACCTCAAGGAACTAAGAGGACACAAAGCTTTACTGTATCTGTAAA	8295	9316	ATCATGATTCCTGTAGGTAAAGTACAGAGCAGCTCTAGATGCTTTCTTCTCTATGAA	9375
QY	8461	ACGCTTATCATGAATCTAAATTTATATTTGGTGTGATCGGTTGTCAGAAATTTGGTACCAT	8520	9541	AGGAGCTCTGTATGACACATGTGCACACACACACACACACACACACACACACACACACAC	9600
DB	8296	ACGCTTATCATGAATCTAAATTTATATTTGGTGTGATCGGTTGTCAGAAATTTGGTACCAT	8355	9376	AGGAGCTCTGTATGACACATGTGCACACACACACACACACACACACACACACACACACAC	9435
QY	8521	GGGCGCTGCGTTGGCATCTTTGCAAAAGTGAGGAGAGCTATTTGATGAGTATGCTCTCCA	8580	9601	GTTTCATGTTAGTATTAATTAAGCTTGCATAAAGGTTGGGCTTAAGTGGTCTTGGGCTAC	9660
DB	8356	GGGCGCTGCGTTGGCATCTTTGCAAAAGTGAGGAGAGCTATTTGATGAGTATGCTCTCCA	8415	9436	GTTTCATGTTAGTATTAATTAAGCTTGCATAAAGGTTGGGCTTAAGTGGTCTTGGGCTAC	9495
				9661	AGACTCTGTTGCTTGAATATAACAGTACAAATTTGTCAATTTACTCTGCACCCAGGCTAAAG	9720

Db	541	TTTTTATAGATGGGATGACGTGGCCAGAGGTCTGTCGGGTGTACTGTGAGAGTATGAAGGAG	600
Qy	859	TACCATCAGCTTCTTCTTACCAAGAGGAGGAGACTACCCATATGGACAGTAGAGAAC	918
Db	601	TACCATCAGCTTCTTCTTACCAAGAGGAGGAGACTACCCATATGGACAGTAGAGAAC	660
Qy	919	AAGATCAAAAGTTCTACAGTTTCTAGTCGATCAGTTTCTTACACAAATATTTGCTCGAGAG	978
Db	661	AAGATCAAAAGTTCTACAGTTTCTAGTCGATCAGTTTCTTACACAAATATTTGCTCGAGAG	720
Qy	979	GAATTCGATGCTCAAGGGGTGATACAGATGATGACCAATGTAGGGTTTGTGCACAACTT	1038
Db	721	GAATTCGATGCTCAAGGGGTGATACAGATGATGACCAATGTAGGGTTTGTGCACAACTT	780
Qy	1039	GGGGATTGCTTGGCTGTGAGACATGTTCCAGCAGTATACCATTTGGAATGTGTGAAGCCA	1098
Db	781	GGGGATTGCTTGGCTGTGAGACATGTTCCAGCAGTATACCATTTGGAATGTGTGAAGCCA	840
Qy	1099	CTCTTCAGAGGTGCCAGAGGAGGAGTGGCAGTGTGAAGTCTGTGTAGCACACAAGGTG	1158
Db	841	CTCTTCAGAGGTGCCAGAGGAGGAGTGGCAGTGTGAAGTCTGTGTAGCACACAAGGTG	900
Qy	1159	CCTGCTGTGACTGACTGTGTGTGAAATCCAAAAAATAAACCATATATTCGACATGAA	1218
Db	901	CCTGCTGTGACTGACTGTGTGTGAAATCCAAAAAATAAACCATATATTCGACATGAA	960
Qy	1219	CCTATTGGATGATGATAGAAAGTCGGGAAATCTGGTCTTGAACCGAAGACTCATAATA	1278
Db	961	CCTATTGGATGATGATAGAAAGTCGGGAAATCTGGTCTTGAACCGAAGACTCATAATA	1020
Qy	1279	GAAGAAGATACAGAAAAATGAAATGAAAGAAAAATTTGGTATTACAGCACAAAAGTCCAA	1338
Db	1021	GAAGAAGATACAGAAAAATGAAATGAAAGAAAAATTTGGTATTACAGCACAAAAGTCCAA	1080
Qy	1339	CTTGACAAATTAATTGACTGTCTAGACAAAGATTTTGGGAAGCAGAACTCTCGCAAAAT	1398
Db	1081	CTTGACAAATTAATTGACTGTCTAGACAAAGATTTTGGGAAGCAGAACTCTCGCAAAAT	1140
Qy	1399	CTAGNAGAAATCCGTGAAGAAATCCCGACACATGGACATACTGAAGACCTGACCAAT	1458
Db	1141	CTAGNAGAAATCCGTGAAGAAATCCCGACACATGGACATACTGAAGACCTGACCAAT	1200
Qy	1459	AAGGCTCGGGCAGTAAACAAATCCTTTCTGGCGCAGCTAATGAAGAAATTTTGGAAATCC	1518
Db	1201	AAGGCTCGGGCAGTAAACAAATCCTTTCTGGCGCAGCTAATGAAGAAATTTTGGAAATCC	1260
Qy	1519	ATAAGAGCCAAAAGGGAGACATTTGATTAATGTTAAAAGCCCGACAGAAACAGAAAAGAC	1578
Db	1261	ATAAGAGCCAAAAGGGAGACATTTGATTAATGTTAAAAGCCCGACAGAAACAGAAAAGAC	1320
Qy	1579	AAGAATGAGCTGAGAATGACTCTAAAGATGCTGAGAAAAACAGAGAAGAAATTTGAAGAC	1638
Db	1321	AAGAATGAGCTGAGAATGACTCTAAAGATGCTGAGAAAAACAGAGAAGAAATTTGAAGAC	1380
Qy	1639	CAGTCCCTTGAAGAGACAGTGCAGACAAAACACAGATGATACCCCTGACAGAGGAAAA	1698
Db	1381	CAGTCCCTTGAAGAGACAGTGCAGACAAAACACAGATGATACCCCTGACAGAGGAAAA	1440
Qy	1699	TCTGAGGAGCCCAACAGAAAGTTGGGGATAAAGTAACTCTGTCTAGCAAAATCTTTGGCCAC	1758
Db	1441	TCTGAGGAGCCCAACAGAAAGTTGGGGATAAAGTAACTCTGTCTAGCAAAATCTTTGGCCAC	1500
Qy	1759	AACACAAACAAATGCAACTTCAGAGAGACTAGTCCCTCTGAAGGGAGAGGCCCTGTGGGG	1818
Db	1501	AACACAAACAAATGCAACTTCAGAGAGACTAGTCCCTCTGAAGGGAGAGGCCCTGTGGGG	1560
Qy	1819	TGTCTCTCAGAAACCCCGATAGCAGACATGCGCAGAGAAAGGTGGCATCTGAGCTC	1878
Db	1561	TGTCTCTCAGAAACCCCGATAGCAGACATGCGCAGAGAAAGGTGGCATCTGAGCTC	1620
Qy	1879	CCCCAGAGTGTGCCAGAAAGCCTTAACAGACATGTGAGAGCAGTAACACTAGTCTACC	1938
Db		TT	
Db	1621	CCCCAGGATGTGCCAGAAAGAACCTTAACAAAGACATGTGAGAGCAGTAACACTAGTCTACC	1680
Qy	1939	ACTACTCTCATCCAGCCTTAATCTGGAAAAACAGTAACAGCAGCAGTGAACATAATCTTCC	1998
Db	1681	ACTACTCTCATCCAGCCTTAATCTGGAAAAACAGTAACAGCAGCAGTGAACATAATCTTCC	1740
Qy	1999	CAGAGTGAATCTGCTTAAGGCAGCTGATCCTGAAATTTGAGAGAAAGAGATCTCATACA	2058
Db	1741	CAGAGTGAATCTGCTTAAGGCAGCTGATCCTGAAATTTGAGAGAAAGAGATCTCATACA	1800
Qy	2059	CCTGTCTCTATTACAGAAAGAGATAGTGTGATTTTCACATCGGAGAACTCCACCGGGAG	2118
Db	1801	CCTGTCTCTATTACAGAAAGAGATAGTGTGATTTTCAMATCGGAGAACTCCACCGGGAG	1860
Qy	2119	CTAAGTGAATCTCTGGAGCTGGAAAAGGAGCATCTGGCTCAACTCGAATCATCACAGA	2178
Db	1861	CTAAGTGAATCTCTGGAGCTGGAAAAGGAGCATCTGGCTCAACTCGAATCATCACAGA	1920
Qy	2179	TTGGGGAATCCAGATAGCAAACTTAGTCAGCTGAAGAGCCAGCAGTGGCAGCCGCTCA	2238
Db	1921	TTGGGGAATCCAGATAGCAAACTTAGTCAGCTGAAGAGCCAGCAGTGGCAGCCGCTCA	1980
Qy	2239	CATGAAGCAAAATAAATTTAAGGAGGCAAGAGAGTACTGGTGTAGTTAACTCTCAAGA	2298
Db	1981	CATGAAGCAAAATAAATTTAAGGAGGCAAGAGAGTACTGGTGTAGTTAACTCTCAAGA	2040
Qy	2299	GAAATTTCCGGTTGAGCACCAAAAAGGAAGTGTATCATGAAAGGAAATATCAACAATAT	2358
Db	2041	GAAATTTCCGGTTGAGCACCAAAAAGGAAGTGTATCATGAAAGGAAATATCAACAATAT	2100
Qy	2359	TTTAAATTTGGTCAAGAGGGAAGTATCGCTCTACCAACAATCAATCTCCACCAATTCA	2418
Db	2101	TTTAAATTTGGTCAAGAGGGAAGTATCGCTCTACCAACAATCAATCTCCACCAATTCA	2160
Qy	2419	TTTGCTTTGAATAAGCAACCCAGCACAGAGAACCATATGATAAGAGAGGCAATCTTGCAAT	2478
Db	2161	TTTGCTTTGAATAAGCAACCCAGCACAGAGAACCATATGATAAGAGAGGCAATCTTGCAAT	2220
Qy	2479	AACTTCTGTCTGACTCCAGCAGGAGATTCATATGAACGGTTCCTCTCCATGGTCCAAA	2538
Db	2221	AACTTCTGTCTGACTCCAGCAGGAGATTCATATGAACGGTTCCTCTCCATGGTCCAAA	2280
Qy	2539	GTTCCTTACCATACTACTCTGAGACTGACTATCACCCCAATTTAGAAAACAACATCCCTTCA	2598
Db	2281	GTTCCTTACCATACTACTCTGAGACTGACTATCACCCCAATTTAGAAAACAACATCCCTTCA	2340
Qy	2599	TCCTTTTTCATCCCAACTGGGCATCACATAGGGCAAAATTTGGAGTGTGAGTTAAACCA	2718
Db	2341	TCCTTTTTCATCCCAACTGGGCATCACATAGGGCAAAATTTGGAGTGTGAGTTAAACCA	2460
Qy	2659	TGTAGCAAAACCCAGAGAAATTTGCATTTGGCTTTAGCCATTTTGGAGTGTGAGTTAAACCA	2778
Db	2401	TGTAGCAAAACCCAGAGAAATTTGCATTTGGCTTTAGCCATTTTGGAGTGTGAGTTAAACCA	2520
Qy	2719	GTTGTGATGTACCAATATGCGGAGAAATTTTGGACATACACAGTTTACACCGGATGACA	2838
Db	2461	GTTGTGATGTACCAATATGCGGAGAAATTTTGGACATACACAGTTTACACCGGATGACA	2580
Qy	2779	TCAAATTCAGAGAGAAAGGAGAAAGTCAAAAAAAGAGAGAAAGAGAGAGAGAGAA	2898
Db	2521	TCAAATTCAGAGAGAAAGGAGAAAGTCAAAAAAAGAGAGAAAGAGAGAGAGAGAA	2640
Qy	2839	GAACAGATGACAGCAAGGCATGGTAAAAATACACATTTCCAGTTTAAGCATCAGTTTGG	2958
Db	2581	GAACAGATGACAGCAAGGCATGGTAAAAATACACATTTCCAGTTTAAGCATCAGTTTGG	2700
Qy	2899	AAACAAAAGGTGAAGAGTACAGAGTACAGAGTATGTTGGTGTGAGCTGGATTAGTAAA	3018
Db	2641	AAACAAAAGGTGAAGAGTACAGAGTACAGAGTATGTTGGTGTGAGCTGGATTAGTAAA	
Qy	2959	ACTCATGTTTATAGTTTGTCTTAAATTTGCCAGGCAATACTAATGTGAATACAGAAAG	
Db	2701	ACTCATGTTTATAGTTTGTCTTAAATTTGCCAGGCAATACTAATGTGAATACAGAAAG	

Qy	3019	TCGTTACGAAGAACCAAAAATAATATGGATGAAAAATATGCGATGAGTCAGATATAAAGAAA	3078
Db	2761	TCGTTAGAAGGAACCAAAAATAATATGGATGAAAAATATGGATGAGTCAGATATAAAGAAA	2820
Qy	3079	TGTTTCAGCAAGTCCAAAAAATAAATAATAGAGCCTGATTCTGAAAAAGATGAGGTAAAA	3138
Db	2821	TGTTTCAGCAAGTCCAAAAAATAAATAATAGAGCCTGATTCTGAAAAAGATGAGGTAAAA	2880
Qy	3139	GGTTTCAGATGCTGCAAAAGGAGCAGACCAAAAATGAAATGGATATCTCAAAAGATTACTGAG	3198
Db	2881	GGTTTCAGATGCTGCAAAAGGAGCAGACCAAAAATGAAATGGATATCTCAAAAGATTACTGAG	2940
Qy	3199	AAGAAGACCAGAAGTGTGAAGGAGCTTTAGATTCGACAGTGATATAACCCCTGCAAGGAA	3258
Db	2941	AAGAAGACCAGAAGTGTGAAGGAGCTTTAGATTCGACAGTGATATAACCCCTGCAAGGAA	3000
Qy	3259	GAACCAATGGAAGTAGACAGATGACATGAAAAACAGAGTCACATGTAAATTTGTCAGGAGAGT	3318
Db	3001	GAACCAATGGAAGTAGACAGATGACATGAAAAACAGAGTCACATGTAAATTTGTCAGGAGAGT	3060
Qy	3319	TCTCAAGTAGATGTGGTCAATGTTAGTGAAGGGTTTTCATCTAAGGACTAGTTACAAAAAG	3378
Db	3061	TCTCAAGTAGATGTGGTCAATGTTAGTGAAGGGTTTTCATCTAAGGACTAGTTACAAAAAG	3120
Qy	3379	AAAAAATAATCACCRAACTAGATGGACTCTTTCGAAAGGAGAAATTAACAGATTACACTG	3438
Db	3121	AAAAAATAATCACCRAACTAGATGGACTCTTTCGAAAGGAGAAATTAACAGATTACACTG	3180
Qy	3439	GAAGAAAAACAGGAGACTCGAAAAAATCAAGTTGGAGGGTGGAAATTAAGGGTATAGGAAG	3498
Db	3181	GAAGAAAAACAGGAGACTCGAAAAAATCAAGTTGGAGGGTGGAAATTAAGGGTATAGGAAG	3240
Qy	3499	ACTTCTACAAATTTCTCAAAAAATCTCTCTGTAATCACCAGTATAAGAAAGCAAAAGAA	3558
Db	3241	ACTTCTACAAATTTCTCAAAAAATCTCTCTGTAATCACCAGTATAAGAAAGCAAAAGAA	3300
Qy	3559	GGGTGTCAGAGTGACTCGATGACAGACAGACAGAGCCCAATGCAATATATGATCAACCT	3618
Db	3301	GGGTGTCAGAGTGACTCGATGACAGACAGACAGAGCCCAATGCAATATATGATCAACCT	3360
Qy	3619	GAGGACTTGATTTCAGGGATGTTCCAAAGTGATTCTCCTCAGTTCTTAGAATGAGTGATCCCT	3678
Db	3361	GAGGACTTGATTTCAGGGATGTTCCAAAGTGATTCTCCTCAGTTCTTAGAATGAGTGATCCCT	3420
Qy	3679	AGTCATACACAAACAAACCTTTATCCAAAAGATCGAGTGTTAGATGATGTCCTCATTCGG	3738
Db	3421	AGTCATACACAAACAAACCTTTATCCAAAAGATCGAGTGTTAGATGATGTCCTCATTCGG	3480
Qy	3739	AGCCAGAAACAAAATGTCGGAACAAAATTCCAATTGAAAATGACATAGAGAAAAAGTC	3798
Db	3481	AGCCAGAAACAAAATGTCGGAACAAAATTCCAATTGAAAATGACATAGAGAAAAAGTC	3540
Qy	3799	TCTGACCTTGCCAGTAGAGGCCAGGACCCACCTAAAGAGTAAAAACCAAGGAAATGATTTT	3858
Db	3541	TCTGACCTTGCCAGTAGAGGCCAGGACCCACCTAAAGAGTAAAAACCAAGGAAATGATTTT	3600
Qy	3859	TTCAATCGATGACTCTAAACTAGCCAGTGCAAGATGATATTGGTACTTTGATCTGTAGAAC	3918
Db	3601	TTCAATCGATGACTCTAAACTAGCCAGTGCAAGATGATATTGGTACTTTGATCTGTAGAAC	3660
Qy	3919	AAAAAACCGCTCATACAGGAGGAAAGTGACACCAATGTTTCTTCTTCCAGAGCTGTTTA	3978
Db	3661	AAAAAACCGCTCATACAGGAGGAAAGTGACACCAATGTTTCTTCTTCCAGAGCTGTTTA	3720
Qy	3979	CATTTCATCAGTGCTTAAAGTACCATGACAGAGATGCCACCTCTGTCAAGAGCAATG	4038
Db	3721	CATTTCATCAGTGCTTAAAGTACCATGACAGAGATGCCACCTCTGTCAAGAGCAATG	3780
Qy	4039	GACTTTGAAGGAAAACTGGGATGTGACTCTGTAATCTAATAGCACTTTGGAAAAATAGTCT	4098
Db	3781	GACTTTGAAGGAAAACTGGGATGTGACTCTGTAATCTAATAGCACTTTGGAAAAATAGTCT	3840

QY	4099	GATACCGTGTCTATTTCAGGATAGCAGTGAAGAGATATGATTTCTTCAGAAATACCAATGAA	4155
DB	3841	GATACCGTGTCTATTTCAGGATAGCAGTGAAGAGATATGATTTCTTCAGAAATACCAATGAA	3900
QY	4159	AGCAATTTCTGAACAGTTTCAGAACTTCGAGAACACAGATGTTTGAAGCTTTGGAGCGGTTAAAG	4218
DB	3901	AGCAATTTCTGAACAGTTTCAGAACTTCGAGAACACAGATGTTTGAAGCTTTGGAGCGGTTAAAG	3960
QY	4219	TGTGAGTTGGTTCTTGGTGTAGTCCACTGGAACACTGTGAGGACAGGCTGCCGGTCAAGGGG	4278
DB	3961	TGTGAGTTGGTTCTTGGTGTAGTCCACTGGAACACTGTGAGGACAGGCTGCCGGTCAAGGGG	4020
QY	4279	ACTGAAGCAAAATGGTAAAAAACCAAGTCACGACAGAAATTTAGAGGAGAGACAGATTAAAT	4338
DB	4021	ACTGAAGCAAAATGGTAAAAAACCAAGTCACGACAGAAATTTAGAGGAGAGACAGATTAAAT	4080
QY	4339	AAATGTAGTGATCAAAATAAGAGCTTAAAAAATACCACTGACAAAAAGAAATATGAAAAATCGA	4398
DB	4081	AAATGTAGTGATCAAAATAAGAGCTTAAAAAATACCACTGACAAAAAGAAATATGAAAAATCGA	4140
QY	4399	GAGTCTGAAAAAGAAAGGACAGAGAACAGTACATTTTCAATAAATGGAAAGATAATAAA	4458
DB	4141	GAGTCTGAAAAAGAAAGGACAGAGAACAGTACATTTTCAATAAATGGAAAGATAATAAA	4200
QY	4459	CCCAAAATATATTTGAAAGGTGAATGCTTGAAGAGAAATTTCTGAGAGTAGAGTAGTAAGT	4518
DB	4201	CCCAAAATATATTTGAAAGGTGAATGCTTGAAGAGAAATTTCTGAGAGTAGAGTAGTAAGT	4260
QY	4519	GGTAATGTTGAACCAAGGTTTAAATAATAAATAAATAATCCCTCAGAGATGATATATAA	4578
DB	4261	GGTAATGTTGAACCAAGGTTTAAATAATAAATAAATAATCCCTCAGAGATGATATATAA	4320
QY	4579	TCATTGACTGTTAAAGAAATCTGCTATTAAGGCCATTCATTAAATGGTGTATGTCATCATCGAA	4638
DB	4321	TCATTGACTGTTAAAGAAATCTGCTATTAAGGCCATTCATTAAATGGTGTATGTCATCATCGAA	4380
QY	4639	GATTTTAAATGAAGAAACAGCTCGGAACAAAATTCGCATTTGCTGTAGTTCTTCAGATGCT	4698
DB	4381	GATTTTAAATGAAGAAACAGCTCGGAACAAAATTCGCATTTGCTGTAGTTCTTCAGATGCT	4440
QY	4699	GAAGTAACATACCGAGATAGCCTTTGAGACCCCTGCCATCAACAAAGAGCTGACACTACA	4758
DB	4441	GAAGTAACATACCGAGATAGCCTTTGAGACCCCTGCCATCAACAAAGAGCTGACACTACA	4500
QY	4759	CAGACGACCACACCCCTCAGCATCTTGTCCAGAAAGCAATTCAGTAAATCAGGTAGAAGAT	4818
DB	4501	CAGACGACCACACCCCTCAGCATCTTGTCCAGAAAGCAATTCAGTAAATCAGGTAGAAGAT	4560
QY	4819	ATGAAATATAGAAACCTCAGAAAGTTAAGAAAGTTACTTTCATCATCTATTTACTTCTGAAGAG	4878
DB	4561	ATGAAATATAGAAACCTCAGAAAGTTAAGAAAGTTACTTTCATCATCTATTTACTTCTGAAGAG	4620
QY	4879	GAATCTAATCTCAGTAATGACTTTATTTGATGAAATGGTCTGCCCATCAACAAAAATGAA	4938
DB	4621	GAATCTAATCTCAGTAATGACTTTATTTGATGAAATGGTCTGCCCATCAACAAAAATGAA	4680
QY	4939	AATGTCATATGGAGAAATCTAAAGAAAAACCGTCATCACAGAACTCACCGATGACCTCC	4998
DB	4681	AATGTCATATGGAGAAATCTAAAGAAAAACCGTCATCACAGAACTCACCGATGACCTCC	4740
QY	4999	ACAGTGGCCACAGAAATCAAAAACCTGTGATCAAGGTAGAAAAAGCGGATTAAGCAAACTGTG	5058
DB	4741	ACAGTGGCCACAGAAATCAAAAACCTGTGATCAAGGTAGAAAAAGCGGATTAAGCAAACTGTG	4800
QY	5059	GTTTCTTCCACAGAAAAATTTGCAAAAATCCACTGTACAAACCACTACACAGTGCACC	5118
DB	4801	GTTTCTTCCACAGAAAAATTTGCAAAAATCCACTGTGTCTCAACCACTACACAGTGCACC	4860
QY	5119	AAGCTTTCCACACCCCTCCACAGGCGGAGTGTGGACATCATCTCTCTAAAGGAGCAGAGC	5178
DB	4861	AAGCTTTCCACACCCCTCCACAGGCGGAGTGTGGACATCATCTCTCTCTAAAGGAGCAGAGC	4920
QY	5179	AAAAACCGTGTGACCCACAGAGTGACAGACTCCCTGACCACCAAGGAGGACCACTGGTT	5238

QY	7399	CAACCCAGGTACAGTCTTCAACTCAAACTCTTTTCATCAGGACAAACTTTTAATCAAGTT	7458
Db	7141	CAACCC-----	7146
QY	7459	AGTGTTCATCCCATCCGTCCTCAGCTACAAATACAGCAGCCACAGCCCAAGTCATT	7518
Db	7147	-----	7146
QY	7519	GCTGTGCCTCAGCTGCAACAACAAGTCCAGGTTCTCTCTCAGATCCAGTCACAGGTTGTG	7578
Db	7147	-----	7146
QY	7579	GCTCAGATACAGGCTCAGCAAAAGTGGTGTGCCCCAGCAAAATCAAACTCCAGTTACCTATC	7638
Db	7147	-----	7146
QY	7639	CAAAATTCAGCAAGCAGTCTGTGCAGACTCACCAGATTCAGAAATGTTTACAGTCGAG	7698
Db	7147	-----	7146
QY	7699	GCAGCCAGTGTGCAAGAGCAGTTGCAAAAGGTTTCAGAACTCAGGGATCAGCAGCAAAAG	7758
Db	7147	-----	7146
QY	7759	AGAAACAGCAACAGATAGAAATTAACGTGAACACACCCCTCAAGCTTCTAATCAAAAGTT	7818
Db	7147	-----	7146
QY	7819	GAATCATTCAGAAACAGTGTGATGAGCATAATGCTGTAATAGAACATTTAAACACAG	7878
Db	7147	-----CAGTGTGATGAGCATAATGCTGTAATAGAACATTTAAACACAG	7191
QY	7879	AAAAAGCATGACTCCAGCTGAAAGAGAGAGAAATCAAAGAAATGATTGCTGTAAACAG	7938
Db	7192	AAAAAGCATGACTCCAGCTGAAGAGAGAGAGATCAAAGAAATGATTGCTGTAAACAG	7251
QY	7939	GTGATGAAGTATATTTTGGATAAGATAGATAAAGAGAAACAGCGCAACAAAAACGG	7998
Db	7252	GTGATGAAGTATATTTTGGATAAGATAGATAAAGAGAAACAGCGCAACAAAAACGG	7311
QY	7999	AGCGTGAAGAGTGTGGAGCAGAAAGTACGACAGAGATGCCACTAAGCTGTCACT	8058
Db	7312	AGCGTGAAGAGTGTGGAGCAGAAAGTACGACAGAGATGCCACTAAGCTGTCACT	7371
QY	8059	CTGCTCTTCAAGCAAAAGAGCAGCTCAGAGCCGAGATCCTGAAGAGAGAGACCTCCTG	8118
Db	7372	CTGCTCTTCAAGCAAAAGAGCAGCTCAGAGCCGAGATCCTGAAGAGAGAGACCTCCTG	7431
QY	8119	GACAAAGATCTGCAAAATTTGAAGTGCAGGAAGAGCTGAAGAGAGACCTGAAAATTAAGAA	8178
Db	7432	GACAAAGATCTGCAAAATTTGAAGTGCAGGAAGAGCTGAAGAGAGACCTGAAAATTAAGAA	7491
QY	8179	GAAGAAGACCTGATGCACTGGCTCAGGCCACAGCAGTAGCTGCACCTGCCCCCAAGTG	8238
Db	7492	GAAGAAGACCTGATGCACTGGCTCAGGCCACAGCAGTAGCTGCACCTGCCCCCAAGTG	7551
QY	8239	ACACCACTTCTCCAGCCCTCCAGCCCTCCACCTTCCACCTCCACCTCCACCTGGTGTG	8298
Db	7552	ACACCACTTCTCCAGCCCTCCAGCCCTCCACCTTCCACCTCCACCTCCACCTGGTGTG	7611
QY	8299	CAACACAGGCTTCTGTCCAGCCCTCCACCTTACCTGTGCTTCCAGAGAGGAGCGG	8358
Db	7612	CAACACAGGCTTCTGTCCAGCCCTCCACCTTACCTGTGCTTCCAGAGAGGAGCGG	7671
QY	8359	GAAGAGAAAGACTCCAGCTCAAGTCCAGAAAAAGAAATGATCTCTACTACCTCA	8418
Db	7672	GAAGAGAAAGACTCCAGCTCAAGTCCAGAAAAAGAAATGATCTCTACTACCTCA	7731
QY	8419	AAGAAACTAAGAGGACACAAAGCTTTACTGTATCTGTAACCCCTTTATGATGAATCT	8478
Db	7732	AAGAAACTAAGAGGACACAAAGCTTTACTGTATCTGTAACCCCTTTATGATGAATCT	7791
QY	8479	AAATTTATATGGCTGTGATCGGTGTGATCAGAAATTTGGTACCATGGCGCTTGGCATC	8538
Db	7792	AAATTTATATTTGGCTGTGATCGGTGTGATCAGAAATTTGGTACCATGGCGCTTGGCATC	7851
QY	8539	TTGCAAGTGAAGCAGAGCTCAATGATGATGATGCTGTCCACAGTGCAGAGTCAACAGAG	8598
Db	7852	TTGCAAGTGAAGCAGAGCTCAATGATGATGATGCTGTCCACAGTGCAGAGTCAACAGAG	7911
QY	8599	GATGCCATGACAGTGTCTCACGCCACTACAGAGAGAGATTATCAGGGGTTGAAGAGGTTG	8658
Db	7912	GATGCCATGACAGTGTCTCACGCCACTACAGAGAGAGATTATCAGGGGTTGAAGAGGTTG	7971
QY	8659	CTCCGTTCTTACAGGCCCATTAAGATGCGCTTCCCTTGTGAACCATGACCCCTAAT	8718
Db	7972	CTCCGTTCTTACAGGCCCATTAAGATGCGCTTCCCTTGTGAACCATGACCCCTAAT	8031
QY	8719	GATGCACAGATTAATATGTTGTTTAAAGAACCTATGAGCTTGGCCACCATGGAAGAA	8778
Db	8032	GATGCACAGATTAATATGTTGTTTAAAGAACCTATGAGCTTGGCCACCATGGAAGAA	8091
QY	8779	AGAGTACAAAGACGATATTATGAAAAGCTGACGGAATTTGTGCAGATATGACCAAAAT	8838
Db	8092	AGAGTACAAAGACGATATTATGAAAAGCTGACGGAATTTGTGCAGATATGACCAAAAT	8151
QY	8839	TTTGATACTGCTGTTACTACAAATCAAGTACTGCCCTTCCCTTACAGTGTGCAGAGTT	8898
Db	8152	TTTGATACTGCTGTTACTACAAATCAAGTACTGCCCTTCCCTTACAGTGTGCAGAGTT	8211
QY	8899	CTCGAATCATCTTTGTACAGAAATTTGAAGCTTCAAGCTAGCAGGTCTCATACAAAC	8958
Db	8212	CTCGAATCATCTTTGTACAGAAATTTGAAGCTTCAAGCTAGCAGGTCTCATACAAAC	8271
QY	8959	AAACTGCAGCTACAGCTTCTTAA	8982
Db	8272	AAACTGCAGCTACAGCTTCTTAA	8295
RESULT 3			
AK025776			
LOCUS	AK025776	3419 bp	linear
DEFINITION	Homo sapiens cDNA: FLJ22123 fis, clone HEP19337, highly similar to AB032251 Homo sapiens BPTF mRNA for bromodomain PHD finger transcription factor.		
ACCESSION	AK025776		
VERSION	AK025776.1	GI:10438394	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (sites) Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	2 (bases 1 to 3419)		
AUTHORS	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (2893009-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)		
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).		
FEATURES	Location/Qualifiers		

	source						
		1..3419					
		/organism="Homo sapiens"					
		/mol_type="mRNA"					
		/db_xref="taxon:9606"					
		/clone="HEP19337"					
		/cell_line="HepG2"					
		/cell_type="hepatoma"					
		/clone_lib="HEP"					
		/note="cloning vector pME18SFL3"					
	misc_feature	1..3419					
		/note="highly similar to AB032251 Homo sapiens BPTF mRNA for bromodomain PHD finger transcription factor"					
	BASE COUNT	1182 a	832 c	709 g	696 t		
	ORIGIN						
		Query Match	30.7%;	Score 3033;	DB 9;	Length 3419;	
		Best Local Similarity	99.6%;	Pred. No. 0;			
		Matches 3062; Conservative	0;	Mismatches 10;	Indels 3;	Gaps 2;	
QY	4337	ATAAATGTAGTGATCAAAATAAAGCTAAAAAATACCACTGACAAAAGAATAATGAAATC	4396				
Db	1	ATAAATGTAGTGATCAAAATAAAGCTAAAAAATACCACTGACAAAAGAATAATGAAATC	60				
QY	4397	GAGAGTCTGAAGAAGGACGAGAGAACAAAGTACATTTCAAATTAATGAAAAGATAATA	4456				
Db	61	GAGAGTCGGAAGAAGGACGAGAGAACAAAGTACATTTCAAATTAATGAAAAGATAATA	120				
QY	4457	AACCCAAAATATATTGAAAGGTGAATGCTTGAAGAAATTTCTGAGACTAGAGTAGTAA	4516				
Db	121	AACCCAAAATATATTGAAAGGTGAATGCTTGAAGAAATTTCTGAGACTAGAGTAGTAA	180				
QY	4517	GTGGTAATGTTGAACCAAAAGTTTAATATAATAATAAATCCCTCGAAGATGATTTA	4576				
Db	181	GTGGTAATGTTGAACCAAAAGTTTAATATAATAATAAATCCCTCGAAGATGATTTA	240				
QY	4577	AATCATTTGACTGTTAAAGATCTGCTATAAGGCCATTCATTAATGTGTGATGCATCATG	4636				
Db	241	AATCATTTGACTGTTAAAGATCTGCTATAAGGCCATTCATTAATGTGTGATGCATCATG	300				
QY	4637	AAGATTTTAATGAAGAACAACAGCTCCGGAACAAAAATCGCATTTGCTGAGTTCTTCAGATG	4696				
Db	301	AAGATTTTAATGAAGAACAACAGCTCCGGAACAAAAATCGCATTTGCTGAGTTCTTCAGATG	360				
QY	4697	CTGAAGGTAACCTACCGAGATAGCGTTGAGACCCTGCCATCAACCAAGAGCTTGACAGTA	4756				
Db	361	CTGAAGGTAACCTACCGAGATAGCGTTGAGACCCTGCCATCAACCAAGAGCTTGACAGTA	420				
QY	4757	CACAGACGACACACCCCTCAGCATCTTGTCCAGAAAGCAATTCAGTTAATCAGGTAGAAG	4816				
Db	421	CACAGACGACACACCCCTCAGCATCTTGTCCAGAAAGCAATTCAGTTAATCAGGTAGAAG	480				
QY	4817	ATATGGAAATAGAAACCTCAGAAGTTAAGAAAGTTACTTCATCACCTTATTACTTCTGAAG	4876				
Db	481	ATATGGAAATAGAAACCTCAGAAGTTAAGAAAGTTACTTCATCACCTTATTACTTCTGAAG	540				
QY	4877	AGGAATCTAATCTCAGTAATGACHTTTATTGATGAAATGCTCTGCCCATCAACAAAAATG	4936				
Db	541	AGGAATCTAATCTCAGTAATGACHTTTATTGATGAAATGCTCTGCCCATCAACAAAAATG	600				
QY	4937	AAAATGTCATGGAAGATCTTAAAAAGAAAAACCGTCATCACAAGAGTCACCACGATGACCT	4996				
Db	601	AAAATGTCATGGAAGATCTTAAAAAGAAAAACCGTCATCACAAGAGTCACCACGATGACCT	660				
QY	4997	CCACAGTGGCCACAGAAATCAAAAACTGTGATCAAGGTAGAAAAAGGCGATAAGCAAACTG	5056				
Db	661	CCACAGTGGCCACAGAAATCAAAAACTGTGATCAAGGTAGAAAAAGGCGATAAGCAAACTG	720				
QY	5057	TGGTTTTCTTCACAGAAAAATTGTGCAAAATCCACTGTGCACAAACCACTACACAGTGA	5116				
Db	721	TGGTTTTCTTCACAGAAAAATTGTGCAAAATCCACTGTGCACAAACCACTACACAGTGA	780				
QY	5117	CCAAGCTTTCCACACCCCTCCACAGGCGCAGTGTGGACATCATCTCTGTAAGAGGACGA	5176				


```
QY 2376 AGGAGATATCGGCTTACACAAATCAATCACTCCACCAATTCATTTGCTTTGTAATAAGCA 2435
|||||
D 2191 AGGGAAGATATCGGCTTACACAAATCAATCACTCCACCAATTCATTTGCTTTGTAATAAGCA 2250
|||||
QY 2436 CCAGCAGAGAGAGACCATATAGAGAGAGGCTTCTTGCACATAAGTCTGTCTGACTCC 2495
|||||
D 2251 CCAGCAGAGAGAGACCATATAGAGAGAGGCTTCTTGCACATAAGTCTGTCTGACTCC 2310
|||||
QY 2496 AGCAGGAGAGTCAAAATGGGAACGGTCTCTGTCATGGTCCAAAGTCTTCAACCATATCTAC 2555
|||||
D 2311 AGCAGGAGAGTCAAAATGGGAACGGTCTCTGTCATGGTCCAAAGTCTTCAACCATATCTAC 2370
|||||
QY 2556 TCTGAGACTGACTATCACCCTATAGAAAACATCCCTTCATCCTTTTTTTCATCCCAA 2615
|||||
D 2371 TCTGAGACTGACTATCACCCTATAGAAAACATCCCTTCATCCTTTTTTTCATCCCAA 2429
|||||
QY 2616 CTGGGCATCATAGGCAATTTGGATCAAGGAGTTCAGATGTGAGCAACCCAGAGA 2675
|||||
D 2430 CTGGGCATCATAGGCAATTTGGATCAAGGAGTTCAGATGTGAGCAACCCAGAGA 2489
|||||
QY 2676 ATTTCATTCGCTTTAGCCATTTGGAGTGTGCGAGTTAAACACCGTGTGATGCTACCAAT 2735
|||||
D 2490 ATTTCATTCGCTTTAGCCATTTGGAGTGTGCGAGTTAAACACCGTGTGATGCTACCAAT 2549
|||||
QY 2736 ATGCGGAGATTTTAGGACATCACCAGTTACCCGGATGACATCAATTTGAAAGAGAGA 2795
|||||
D 2550 ATGCGGAGATTTTAGGACATCACCAGTTACCCGGATGACATCAATTTGAAAGAGAGA 2609
|||||
QY 2796 AAAGGAGAAATCAAAAAAAGAGAGAAACAGGAGAGAGAGAG 2839
|||||
D 2610 AAAGGAGAAATCAAAAAAAGAGAGAAACAGGAGAGAGAGAG 2653
|||||

RESULT 5
AC107377/c
LOCUS Homo sapiens chromosome 17, clone CTD-2314M10, complete-sequence.
DEFINITION AC107377
ACCESSION AC107377
VERSION AC107377.4 GI:20377031
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 88848)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
PRT-01-MAY-2002
Unpublished
Homo sapiens chromosome 17, clone CTD-2314M10
2 (bases 1 to 88848)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., LaRocque,K., Lamazares,R.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 1, 2002 this sequence version replaced gi:20303884.
All repeats were identified using RepeatMasker:
Smith, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L21459
Center clone name: 2314_M_10
----- Location/Qualifiers
1 . 88848
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"

JOURNAL
REFERENCE
AUTHORS
SUBMITTED (19-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 88848)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., LaRocque,K., Lamazares,R.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 88848)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., LaRocque,K., Lamazares,R.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 1, 2002 this sequence version replaced gi:20303884.
All repeats were identified using RepeatMasker:
Smith, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L21459
Center clone name: 2314_M_10
----- Location/Qualifiers
1 . 88848
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"

FEATURES
source
```

repeat_region	/map="17"	
repeat_region	/clone="CTD-2314M10"	
repeat_region	/clone_lib="CITDI Human BAC"	
repeat_region	complement(98..394)	
repeat_region	/rpt_family="AluSc"	
repeat_region	408..446	
repeat_region	/rpt_family="L2"	
repeat_region	complement(1891..2157)	
repeat_region	/rpt_family="AluSx"	
repeat_region	2253..2538	
repeat_region	/rpt_family="L2"	
repeat_region	complement(2618..2816)	
repeat_region	/rpt_family="AluSx"	
repeat_region	2882..3044	
repeat_region	/rpt_family="L2"	
repeat_region	3053..3253	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(4123..4358)	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(4539..4848)	
repeat_region	/rpt_family="AluJb"	
repeat_region	complement(5101..5396)	
repeat_region	/rpt_family="AluSc"	
repeat_region	complement(5449..5754)	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(5927..6211)	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(6398..6702)	
repeat_region	/rpt_family="AluJo"	
repeat_region	complement(6715..7020)	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(7275..7572)	
repeat_region	/rpt_family="AluSx"	
repeat_region	7725..7745	
repeat_region	/rpt_family="AT_rich"	
repeat_region	8488..8513	
repeat_region	/rpt_family="(TTTA)n"	
repeat_region	complement(8517..8666)	
repeat_region	/rpt_family="FRAM"	
repeat_region	8667..8695	
repeat_region	/rpt_family="(TTCA)n"	
repeat_region	complement(8714..9005)	
repeat_region	/rpt_family="AluSc"	
repeat_region	complement(9022..9328)	
repeat_region	/rpt_family="AluSx"	
repeat_region	9492..9785	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(9849..10156)	
repeat_region	/rpt_family="AluJb"	
repeat_region	11807..12100	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(12718..12832)	
repeat_region	/rpt_family="MIR"	
repeat_region	13094..13244	
repeat_region	/rpt_family="FRAM"	
repeat_region	13367..13673	
repeat_region	/rpt_family="AluJ"	
repeat_region	13686..13997	
repeat_region	/rpt_family="AluJo"	
repeat_region	14455..14475	
repeat_region	/rpt_family="AT_rich"	
repeat_region	complement(14478..14608)	
repeat_region	/rpt_family="FLAM_c"	
repeat_region	14917..14938	
repeat_region	/rpt_family="(A)n"	
repeat_region	14946..15215	
repeat_region	/rpt_family="AluJb"	
repeat_region	15576..15735	
repeat_region	/rpt_family="AluSx"	
repeat_region	15736..15982	
repeat_region	/rpt_family="AluSx"	
repeat_region	16211..16518	
repeat_region	/rpt_family="AluSx"	
unsure	16519..16524	
repeat_region	/note="<30 Qual SNGL Region"	
repeat_region	16635..16933	
repeat_region	/rpt_family="AluJb"	
repeat_region	complement(17254..17410)	
repeat_region	/rpt_family="MER58A"	
repeat_region	17422..17547	
repeat_region	/rpt_family="L2"	
repeat_region	complement(17944..18240)	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(18426..18746)	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(18751..19065)	
repeat_region	/rpt_family="AluSx"	
repeat_region	19095..19142	
repeat_region	/rpt_family="AT_rich"	
repeat_region	complement(19477..19574)	
Query Match 23.5%; Score 2316; DB 9; Length 88848;		
Best Local Similarity 99.6%; Pred. No. 0;		
Matches 2322; Conservative 0; Mismatches 10; Indels 0; Gaps 0;		
Qy	3210	AGATGTCAGGAGCTCTAGATTCTGACAGTGATAAACCCTGCAAGGAAGAACCAATGGA 3269
Db	29115	AGATGTGAAGGAGCTCTTAGATTCTGACAGTGATAAACCCTGCAAGGAAGAACCAATGGA 29056
Qy	3270	AGTAGACGATGACATGAACACAGAGTCACATGTAATGTGACGAGAGTTCTCAAGTAGA 3329
Db	29055	AGTAGACGATGACATGAACACAGAGTCACATGTAATGTGACGAGAGTTCTCAAGTAGA 28996
Qy	3330	TGTGGTCAATGTTAGTGAGGGTTTTCATCTAAGGACTAGTTACAAAAGAAAACAAAATC 3389
Db	28995	TGTGGTCAATGTTAGTGAGGGTTTTCATCTAAGGACTAGTTACAAAAGAAAACAAAATC 28936
Qy	3390	ATCCAAACTAGATGGACTTCTTGAAGGAGAAATTAACACAGTTTACACTGGAGAAAACA 3449
Db	28935	ATCCAAACTAGATGGACTTCTTGAAGGAGAAATTAACACAGTTTACACTGGAGAAAACA 28876
Qy	3450	CGGACTCGAAAAATCAAGTTGGAGGTGGAATTAAGGGTATAGGAAGACTTCTACAA 3509
Db	28875	CGGACTCGAAAAATCAAGTTGGAGGTGGAATTAAGGGTATAGGAAGACTTCTACAA 28816
Qy	3510	TTCTTCAAAAATCTCTGAATCACCAGTAATAACGAAAAGAAAAGGGGTGTCAGAG 3569
Db	28815	TTCTTCAAAAATCTCTGAATCACCAGTAATAACGAAAAGAAAAGGGGTGTCAGAG 28756
Qy	3570	TGACTCGATGAGACAAGAACAGCCCAATGCAATTAATGATCAACCTGAGACTTGAT 3629
Db	28755	TGACTCGATGAGACAAGAACAGCCCAATGCAATTAATGATCAACCTGAGACTTGAT 28696
Qy	3630	TCAGGGATGTCACAAAGTATTCCTCAGTTCTTAGAATGAGTGATCCTAGTCAATACC 3689
Db	28695	TCAGGGATGTCACAAAGTATTCCTCAGTTCTTAGAATGAGTGATCCTAGTCAATACC 28636
Qy	3690	AAACAAACTTTATCCAAAGATCGAGTGTAGATGATGTCTCCATTCGGAGCCAGAAAC 3749
Db	28635	AAACAAACTTTATCCAAAGATCGAGTGTAGATGATGTCTCCATTCGGAGCCAGAAAC 28576
Qy	3750	AAAATGTCGGAACAAAATTCATTTGAAATGACATAGAGAAAAGTCTGTGACCTGC 3809
Db	28575	AAAATGTCGGAACAAAATTCATTTGAAATGACATAGAGAAAAGTCTGTGACCTGC 28516
Qy	3810	CAGTAGAGCCAGGAAACCCACTAAGAGTAAAAACCAAGAAATGATTTTTTCATCGATGA 3869
Db	28515	CAGTAGAGCCAGGAAACCCAGTAAAGAGTAAAAACCAAGAAATGATTTTTTCATCGATGA 28456
Qy	3870	CTCTAACTAGCCAGTGCAGATGATATTGGTACTTTGATCTGTAGAACAAAACCCGCT 3929
Db	28455	CTCTAACTAGCCAGTGCAGATGATATTGGTACTTTGATCTGTAGAACAAAACCCGCT 28396
Qy	3930	CATACAGGAGGAAAGTGACACCATTTCTTCTCCAGAGTGTCTTTACATTATCATCAGT 3989
Db	28395	CATACAGGAGGAAAGTGACACCATTTCTTCTTCCAGAGTGTCTTTACATTATCATCAGT 28336

Qy	3990	GCCTAAAGGTACCAATGACAGAGATGCCACACCTCTGTCAAGAGCAATGGACTTTGAAGG	4049
Db	28335	GCCTAAAGGTACCAATGACAGAGATGCCACACCTCTGTCAAGAGCAATGGACTTTGAAGG	28276
Qy	4050	AAAACCTGGGATGTGACTCTGAATCTAATAGCACTTTGGAAATAGTCTCTGATACCGTCTC	4109
Db	28275	AAAACCTGGGATGTGACTCTGAATCTAATAGCACTTTGGAAATAGTCTCTGATACCGTCTC	28216
Qy	4110	TATTCAGGATAGCAGTGAAGAAGATATGTTCTCAGAAATGCAATGAAGCATTTCTTGA	4169
Db	28215	TATTCAGGATAGCAGTGAAGAAGATATGTTCTCAGAAATGCAATGAAGCATTTCTTGA	28156
Qy	4170	ACAGTTCAGAACTTCGAGAACACAGATGTTGAACTCTTTGGAGCCGTTAAAGTGTGAGTTGGT	4229
Db	28155	ACAGTTCAGAACTTCGAGAACACAGATGTTGAACTCTTTGGAGCCGTTAAAGTGTGAGTTGGT	28096
Qy	4230	TTCTGGTGGAGTCCACCTGGAAACTGTGAGCACAGGCTGCCGGTCAAGGGGACTGAAGCAAA	4289
Db	28095	TTCTGGTGGAGTCCACCTGGAAACTGTGAGCACAGGCTGCCGGTCAAGGGGACTGAAGCAAA	28036
Qy	4290	TGTTAAAAAACCAAGTCAGCAGAAGAAATTAGAGGAGAGACCAAGTAAATAATGTAAGTGA	4349
Db	28035	TGTTAAAAAACCAAGTCAGCAGAAGAAATTAGAGGAGAGACCAAGTAAATAATGTAAGTGA	27976
Qy	4350	TCAAAATAAGCTTAAAAAATACCACTGACAAAAAGAAATAATGAAATCGAGAGTCTGAAAA	4409
Db	27975	TCAAAATAAGCTTAAAAAATACCACTGACAAAAAGAAATAATGAAATCGAGAGTCTGAAAA	27916
Qy	4410	GAAAGCACAGAGAACCAAGTACATTTCAAAATAATGGAAGAAGATAATAACCCAAAATATA	4469
Db	27915	GAAAGCACAGAGAACCAAGTACATTTCAAAATAATGGAAGAAGATAATAACCCAAAATATA	27856
Qy	4470	TTTGAAGGTGAATGCTTGAAAGAAATTTCTGAGAGTAGAGTAGTAAGTGGTAATGTTGA	4529
Db	27855	TTTGAAGGTGAATGCTTGAAAGAAATTTCTGAGAGTAGAGTAGTAAGTGGTAATGTTGA	27796
Qy	4530	ACCAAAGGTTAATAATATAATAATAATATCCCTGAGAAATGATATTAATCAATTCAGTGT	4589
Db	27795	ACCAAAGGTTAATAATATAATAATAATATCCCTGAGAAATGATATTAATCAATTCAGTGT	27736
Qy	4590	TAAAGAACTGCTATTAAGGCCATTCATTAATGGTGATGTCATCATCGGAAGATTTTAATGA	4649
Db	27735	TAAAGAACTGCTATTAAGGCCATTCATTAATGGTGATGTCATCATCGGAAGATTTTAATGA	27676
Qy	4650	AAGAAACAGCTCCGAAACAAAATCGCATTTGCTGAGTTCTTCAGAGTCTGGAAGTAACTA	4709
Db	27675	AAGAAACAGCTCCGAAACAAAATCGCATTTGCTGAGTTCTTCAGAGTCTGGAAGTAACTA	27616
Qy	4710	CCGAGATAGCTTTGAGACCCCTGCCATCAACCAAAGAGTCTGACAGTACACAGACGACCAC	4769
Db	27615	CCGAGATAGCTTTGAGACCCCTGCCATCAACCAAAGAGTCTGACAGTACACAGACGACCAC	27556
Qy	4770	ACCTTCAGCATCTTTGTCAGAAAGCAATTCAGTTAATCAGGTAGAAGATATGGAATAAGA	4829
Db	27555	ACCTTCAGCATCTTTGTCAGAAAGCAATTCAGTTAATCAGGTAGAAGATATGGAATAAGA	27496
Qy	4830	AACTCAGAAAGTTAAGAAAGTTACTTTCATCACCTATTACTCTGGAAGGAGTAATCTAATCT	4889
Db	27495	AACTCAGAAAGTTAAGAAAGTTACTTTCATCACCTATTACTCTGGAAGGAGTAATCTAATCT	27436
Qy	4890	CAGTAATGACTTTATTGATGAAATGGTCTGCCCATCAACAAAAATGAAATGTCAATGG	4949
Db	27435	CAGTAATGACTTTATTGATGAAATGGTCTGCCCATCAACAAAAATGAAATGTCAATGG	27376
Qy	4950	AGAAATCAAAGAAACCCGTCATCACAGAAAGTCAACCAAGTACCTCCACAGTGGCCAC	5009
Db	27375	AGAAATCAAAGAAACCCGTCATCACAGAAAGTCAACCAAGTACCTCCACAGTGGCCAC	27316
Qy	5010	AGAAATCAAAACTGTGATCAAGGTAGAAAAAGGCCATAAAGCAAACTGTGGTTTCTTCCAC	5069
Db	27315	AGAAATCAAAACTGTGATCAAGGTAGAAAAAGGCCATAAAGCAAACTGTGGTTTCTTCCAC	27256

Qy	5070	AGAAATTGTGCAAAATCCACTCTCAACACCACTACACAGTACCAAGCTTTCCAC	5129
Db	27255	AGAAATTGTGCAAAATCCACTCTCAACACCACTACACAGTACCAAGCTTTCCAC	27196
Qy	5130	ACCTCCACAGCGGCGAGTGTGACATCATCTCTGTAAAGGAGCAGACGAAACCGTGT	5189
Db	27195	ACCTCCACAGCGGCGAGTGTGACATCATCTCTGTAAAGGAGCAGACGAAACCGTGT	27136
Qy	5190	CACCACGACAGTGTGACAGACTCCCTGACCCACCGGAGGACACACTGGTTACATCTATGAC	5249
Db	27135	CACCACGACAGTGTGACAGACTCCCTGACCCACCGGAGGACACACTGGTTACATCTATGAC	27076
Qy	5250	TGTGAGCAAGAGTATTTCCACAGCAGACAAAGTGAACATGTAGAAATTTTCAAGACCAA	5309
Db	27075	TGTGAGCAAGAGTATTTCCACAGCAGACAAAGTGAACATGTAGAAATTTTCAAGACCAA	27016
Qy	5310	GAAGACTCGTTCAGGTACAGCTCGCCATCTATAGAAATTTGTTACCAAGAGCACC	5369
Db	27015	GAAGACTCGTTCAGGTACAGCTCGCCATCTATAGAAATTTGTTACCAAGAGCACC	26956
Qy	5370	GAAGAGCATTTTGTGCTTAATGATGACTTTAAAAAGTTGGCCCGAAAAAGGAGGAAT	5429
Db	26955	GAAGAGCATTTTGTGCTTAATGATGACTTTAAAAAGTTGGCCCGAAAAAGGAGGAAT	26899
Qy	5430	CCGAGAGTCCCTTATTTAATTTACAACTGCTTTGGGATATATGCGCATATCC	5489
Db	26895	CCGAGAGTCCCTTATTTAATTTACAACTGCTTTGGGATATATGCGCATATCC	26836
Qy	5490	TTCTCCTAGACCGACCTTTGGCATCAGTTGGAGGTATAGACTTCACAGTGA	5541
Db	26835	TTCTCCTAGACCGACCTTTGGCATCAGTTGGAGGTATAGACTTTAAATGTGA	26784
RESULT 6			
AC127525/c			
LOCUS	AC127525	108860 bp	DNA linear PRI 30-OCT-2002
DEFINITION	Homo sapiens chromosome 17, clone RP13-551H18, complete sequence.		
ACCESSION	AC127525		
VERSION	AC127525.8 GI:24421744		
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 108860)		
TITLE	Birren,B., Nusbaum,C. and Lander,E.		
JOURNAL	Homo sapiens chromosome 17, clone RP13-551H18		
AUTHORS	Unpublished		
2 (bases 1 to 108860)			
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhaltier,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galaugan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., PhunKhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.			
REFERENCE	Homo sapiens chromosome 17, clone RP13-551H18		
AUTHORS	Unpublished		
2 (bases 1 to 108860)			
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhaltier,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galaugan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., PhunKhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.			
JOURNAL	Submitted (17-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE	3 (bases 1 to 108860)		
AUTHORS	Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhaltier,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galaugan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., PhunKhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.		

Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,
 Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T.,
 Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
 Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (42-OCT-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 108860)
 Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,
 Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T.,
 Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
 Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (30-OCT-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Oct 30, 2002 this sequence version replaced gi:23915491.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RN/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L27577
 Center clone name: 551_H_18

 Location/Qualifiers
 1. .108860
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /Clone="RP13-551H18"
 /Clone_lib="RPC1-13 Human Female BAC"
 complement(1. .61)
 /rpt_family="AluY"
 81. 240
 /rpt_family="MIR"
 241. 537
 /rpt_family="AluSx"
 569. 592
 /rpt_family="AT-rich"
 607. 898
 /rpt_family="AluYb"
 complement(904. .950)
 /rpt_family="MLT1F1"
 complement(951. .1230)
 /rpt_family="AluSg"
 complement(1231. .1569)
 /rpt_family="MLT1F1"
 1570. .1685
 /rpt_family="L1Pb1"
 1686. .1713
 /rpt_family="(TG)n"
 1713. .1743
 /rpt_family="(TA)n"
 1714. .2101
 /rpt_family="L1Pb1"
 complement(2102. .2277)
 /rpt_family="MLT1F1"
 complement(2292. .2576)
 /rpt_family="AluSg"
 2707. .2732
 /rpt_family="AT-rich"
 complement(2795. .3089)
 /rpt_family="AluSx"
 3092. .3121
 /rpt_family="(TA)n"
 complement(3122. .3419)
 /rpt_family="AluSx"
 3448. .3758
 /rpt_family="AluSc"
 3869. .4006
 /rpt_family="FLAM_C"
 4153. .4489
 /rpt_family="AluYb"
 4628. .4932
 /rpt_family="AluY"
 5080. .5390
 /rpt_family="AluY"
 5407. .5701
 /rpt_family="AluSp"
 5743. .6051
 /rpt_family="AluSg"
 6058. .6362
 /rpt_family="AluY"
 6366. .6667
 /rpt_family="AluY"
 6672. .6818
 /rpt_family="L1M3"
 6822. .7143
 /rpt_family="AluSx"
 7238. .7338
 /rpt_family="L1M3"
 7342. .7433
 /rpt_family="(TA)n"
 7460. .7749
 /rpt_family="AluSg"
 7851. .7970
 /rpt_family="L1M3"
 complement(7971. .8272)
 /rpt_family="AluSx"
 8273. .8322
 /rpt_family="L1M3"
 8331. .8394
 /rpt_family="MER21B"
 complement(8412. .8563)
 /rpt_family="AluYb"
 8587. .8618
 /rpt_family="L12"
 complement(8980. .9289)
 /rpt_family="AluSg"
 complement(9866. .10174)
 /rpt_family="AluSg"
 complement(10220. .10548)
 /rpt_family="AluY"
 complement(10594. .10893)

[illegible]

Db	95496	AGAAAATTGCGAAATCCACTGTCACACCCACCCTACACAGTACCAAGCTTTCAC		95437
Qy	5130	ACCCTCACAGGGCGAGTGGACATCATCTCTGTAAAGGACGACGAAAACCGTGGT		5189
Db	95436	ACCCTCACAGGGCGAGTGGACATCATCTCTGTAAAGGACGACGAAAACCGTGGT		95377
Qy	5190	CACCACGACAGTGACAGACCTCCCTGACACCCACCGGAGGACACTGGTTCATCTATGAC		5249
Db	95376	CACCACGACAGTGACAGACCTCCCTGACACCCACCGGAGGACACTGGTTCATCTATGAC		95317
Qy	5250	TGTGAGCAAAAGTATTCCACAGACGACAAAGTGAACACTGATGAAATTTTCAGAGACAAA		5309
Db	95316	TGTGAGCAAAAGTATTCCACAGACGACAAAGTGAACACTGATGAAATTTTCAGAGACAAA		95257
Qy	5310	GAAGACTGTCAGGTACAGTCTGCCATCCTATAGAAAATTTGTACCAAGACGACCAA		5369
Db	95256	GAAGACTGTCAGGTACAGTCTGCCATCCTATAGAAAATTTGTACCAAGACGACCAA		95197
Qy	5370	GAAGAGCATTTTCTTTTGGCTAATGATGACTTAAAAAAGTTGCCCGAAAAGGAGGAAT		5429
Db	95196	GAAGAGCATTTTCTTTTGGCTAATGATGACTTAAAAAAGTTGCCCGAAAAGGAGGAAT		95137
Qy	5430	CCGAGAGTCCCTTATTTTAAATACAAATGCAAAACCTGCTTGGATATATGGCCATATCC		5489
Db	95136	CCGAGAGTCCCTTATTTTAAATACAAATGCAAAACCTGCTTGGATATATGGCCATATCC		95077
Qy	5490	TTCTCCTAGACCGACCTTTGGCATCACTTGGAGGTATAGACTTCAGACACGTA	5541	
Db	95076	TTCTCCTAGACCGACCTTTGGCATCACTTGGAGGTATAGACTTCAGACACGTA	95025	
RESULT 7				
AC134407				
LOCUS	AC134407	199875 bp	DNA	linear
DEFINITION	Homo sapiens chromosome 17, clone Rpl1-855A2, complete sequence.			
ACCESSION	AC134407			
VERSION	AC134407.6	GI:31455710		
KEYWORDS	HTG.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Birren,B., Nusbaum,C. and Lander,E.			
JOURNAL	Homo sapiens chromosome 17, clone Rpl1-855A2			
AUTHORS	Unpublished			
	2 (bases 1 to 199875)			
	Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gordan,S., Guld,S., Graham,L., Grand-Pierre,N., Hages,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mhova,T., Norbu,C., Norman,C.H., O'Connor,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Peterson,K., Phunkhang,P., O'Neill,D., Oliver,J., Peterson,K., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.			
	Direct Submission			
	Submitted (26-SEP-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
	3 (bases 1 to 199875)			
	Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hages,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mhova,T., Norbu,C., Norman,C.H., O'Connor,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Peterson,K., Phunkhang,P., O'Neill,D., Oliver,J., Peterson,K., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.			
	Direct Submission			
	Submitted (21-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
	5 (bases 1 to 199875)			
	Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hages,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mhova,T., Norbu,C., Norman,C.H., O'Connor,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Peterson,K., Phunkhang,P., O'Neill,D., Oliver,J., Peterson,K., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.			
	Direct Submission			
	Submitted (21-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
	5 (bases 1 to 199875)			
	Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hages,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mhova,T., Norbu,C., Norman,C.H., O'Connor,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Peterson,K., Phunkhang,P., O'Neill,D., Oliver,J., Peterson,K., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.			
	Direct Submission			
	Submitted (21-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
	6 (bases 1 to 199875)			
	Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,			


```
|||||
42783 GCCTAAAGTACCAATGACAGAGATGCCACACCTCTGTCAAGAGCAATGGACTTTGAAGG 42842
QY AAAACTGGGATGTGACTCTCTGAATCTAATAGCACTTTGGAAAAATAGTTCTGATACCGTGTG 4109
Db AAAACTGGGATGTGACTCTCTGAATCTAATAGCACTTTGGAAAAATAGTTCTGATACCGTGTG 42902
QY TATTCAGATAGCAGTCAAGAAGATGATTGTTTCAAGATAGCAATGAAAGCAATTTCTGA 4169
Db TATTCAGATAGCAGTCAAGAAGATGATTGTTTCAAGATAGCAATGAAAGCAATTTCTGA 42962
QY ACAGTTTCAGAACTCGGAACAAGATGTTGAAGTCTTGGAGCCGTTAAAGTGTGAGTTGGT 4229
Db ACAGTTTCAGAACTCGGAACAAGATGTTGAAGTCTTGGAGCCGTTAAAGTGTGAGTTGGT 43022
QY TTCTGGTGTGACTGAACTGTGAGGACAGGCTGCCGGTCAAGGGGACTGAAGCAAA 4289
Db TTCTGGTGTGACTGAACTGTGAGGACAGGCTGCCGGTCAAGGGGACTGAAGCAAA 43082
QY TGGTAAAAAACCAAGTCAGCAGAGAAATTTAGAGGAGAGACCAAGTTAAATAAGTGTAGTGA 4349
Db TGGTAAAAAACCAAGTCAGCAGAGAAATTTAGAGGAGAGACCAAGTTAAATAAGTGTAGTGA 43142
QY TCAAAATAAGCTAAAAATACCAGTCGACAAAAAGAAATTAAGAAATCGAGAGTCTGAAAA 4409
Db TCAAAATAAGCTAAAAATACCAGTCGACAAAAAGAAATTAAGAAATCGAGAGTCTGAAAA 43202
QY GAAAGGACAGAGACAGATACATTTCAATTAATGGAAGAGATTAATGAAACCCCAAAATATA 4469
Db GAAAGGACAGAGACAGATACATTTCAATTAATGGAAGAGATTAATGAAACCCCAAAATATA 43262
QY TTTGAAAGGTCGATGCTGAAAGAAATTTCTGAGAGTAGAGTAGTAGTGAAGTGTGATGTTGA 4529
Db TTTGAAAGGTCGATGCTGAAAGAAATTTCTGAGAGTAGAGTAGTAGTGAAGTGTGATGTTGA 43322
QY ACCAAAGGTTAATAATAATAATAATAATCCCTCGAGAAGATTAATAATCAATTCAGTGT 4589
Db ACCAAAGGTTAATAATAATAATAATAATCCCTCGAGAAGATTAATAATCAATTCAGTGT 43382
QY TAAAGAAATCTCTAAGAGCCATTCATTAAATGTTGATGTCATCATGGAAGATTTTAATGA 4649
Db TAAAGAAATCTCTAAGAGCCATTCATTAAATGTTGATGTCATCATGGAAGATTTTAATGA 43442
QY AAGAAACAGCTCCGAAACAAATCCGATTTGCTGAGTCTTCAGATGCTGAAGTAACTA 4709
Db AAGAAACAGCTCCGAAACAAATCCGATTTGCTGAGTCTTCAGATGCTGAAGTAACTA 43502
QY CCGAGATAGCCTTGAGACCCCTGCCATCAACCAAGAGTCTGACAGTACACAGACGACAC 4769
Db CCGAGATAGCCTTGAGACCCCTGCCATCAACCAAGAGTCTGACAGTACACAGACGACAC 43562
QY ACCCTCAGCATCTTGTCCAGAAAGCAATTCAGTTAATCAGGTAGAAGATATGGAATAGA 4829
Db ACCCTCAGCATCTTGTCCAGAAAGCAATTCAGTTAATCAGGTAGAAGATATGGAATAGA 43622
QY AACCTCAGAGTTAAGAAGTTACTTCATCACTATTACTTCTGAAGAGGAATCTAATCT 4889
Db AACCTCAGAGTTAAGAAGTTACTTCATCACTATTACTTCTGAAGAGGAATCTAATCT 43682
QY CAGTAATCAGTTATTGATGAAATGGTCTGCCATCAACAAAAATGAAATGTCAATGG 4949
Db CAGTAATCAGTTATTGATGAAATGGTCTGCCATCAACAAAAATGAAATGTCAATGG 43742
QY AGAATCTAAAAAGAAACCGTCATCACAGAGTCAACACGATGACCTCCACAGTGGCCAC 5009
Db AGAATCTAAAAAGAAACCGTCATCACAGAGTCAACACGATGACCTCCACAGTGGCCAC 43802
QY AGAATCAAAAACCTGTGTGATAGAAAAGGCGATAGCAAACTGTGGTTCTTCCAC 5069
Db AGAATCAAAAACCTGTGTGATAGAAAAGGCGATAGCAAACTGTGGTTCTTCCAC 43862
QY AGAAATTTGTCAAAATCCACTGTTCACACACCACTACAACTGACCAAGTCTTCCAC 5129
Db AGAAATTTGTCAAAATCCACTGTTCACACACCACTACAACTGACCAAGTCTTCCAC
```

```
Db 43863 AGAAAAATTGTCAAAATCCACTGTCTACAAACCCACTACAAACAGTGACCAAGCTTTCCAC 43922
QY 5130 ACCCTCCACAGGGCGAGTGTGGACATCATCTCTGTAAAGGAGCAGACGAAAAACCGTGT 5189
Db 43923 ACCCTCCACAGGGCGAGTGTGGACATCATCTCTGTAAAGGAGCAGACGAAAAACCGTGT 43982
QY 5190 CACCACGACAGTGACAGACTCCCTGACCACCGAGGAGGACACTGGTTACATCTATGAC 5249
Db 43983 CACCACGACAGTGACAGACTCCCTGACCACCGAGGAGGACACTGGTTACATCTATGAC 44042
QY 5250 TGTGAGCAAGAGATTTCCACACGACAGCAAGTGAACCTGATGAAATTTTCAAGACCAA 5309
Db 44043 TGTGAGCAAGAGATTTCCACACGACAGCAAGTGAACCTGATGAAATTTTCAAGACCAA 44102
QY 5310 GAAGACTCGTTCAAGTACAGCTCTGCCATCTCTATAGAAAAATTTGTACCAAGAGCACCAA 5369
Db 44103 GAAGACTCGTTCAAGTACAGCTCTGCCATCTCTATAGAAAAATTTGTACCAAGAGCACCAA 44162
QY 5370 GAAGAGCATTTTGTGCTTAATGATGACTTTAAAAAGTTGGCCCGGAAAAAGAGGAAT 5429
Db 44163 GAAGAGCATTTTGTGCTTAATGATGACTTTAAAAAGTTGGCCCGGAAAAAGAGGAAT 44222
QY 5430 CCAGAGGTCCTTATTTTAAATTTACAATGCAAAACCTGCTTTGGATATATATGGCCATATCC 5489
Db 44223 CCAGAGGTCCTTATTTTAAATTTACAATGCAAAACCTGCTTTGGATATATATGGCCATATCC 44282
QY 5490 TTCTCTAGACCGACCTTTGGCATCACTTTGGAGGTATAGACTTCAGACAGTA 5541
Db 44283 TTCTCTAGACCGACCTTTGGCATCACTTTGGAGGTATAGACTTTAAAAATGTA 44334
```

```
RESULT 8
HSM804688 HSM804688 4074 bp mRNA linear PRI 10-JUL-2002
LOCUS Homo sapiens mRNA; cDNA DKFZp667H1714 (from clone DKFZp667H1714).
DEFINITION AL833375
ACCESSION AL833375.1 GI:21734010
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4074)
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by MediGenomix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project. This clone
(DKFZp667H1714) is available at the RZPD in Berlin. Please contact
the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: cloneerzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
```

```
FEATURES
source
1..4074
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp667H1714"
/tissue_type="lymph node"
/clone_lib="667 (synonym: hlno2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
polya_signal
polya_site
4028..4033
BASE COUNT 1280 a 874 c 788 g 1132 t
ORIGIN
```

```
Query Match 20.3%; Score 1998; DB 9; Length 4074;
Best Local Similarity 81.1%; Pred. No. 4e-310;
Matches 2638; Conservative 0; Mismatches 10; Indels 604; Gaps 3;
```

QY	6777	GACTGTACTCCAGGCCAGCCAGCAGCTAAATGCAAGCTGCAATGCCAAATGGTACTGT	6836
DB	13	GACTGTACTCCAGGCCAGCCAGCAGCTAAATGCAAGCTGCAATGCCAAATGGTACTGT	72
QY	6837	TCAGCGAATCTCTTTACCCCATTTGGCAACAACAGCCACCACAGCCAGCCACCACCAC	6896
DB	73	TCAGCGAATCTCTTTACCCCATTTGGCAACAACAGCCACCACAGCCAGCCACCACCAC	132
QY	6897	CACGTGTTTCCACGACAGCAGCAGTGAAGTGAACAAAGGACAGAGTAAACTGTCAACCCCA	6956
DB	133	CACGTGTTTCCACGACAGCAGCAGTGAAGTGAACAAAGGACAGAGTAAACTGTCAACCCCA	192
QY	6957	GATCGAGGTACATCAAGACAAAACCCCTGCCACAGCTCAGTCATCAAGTGTGGTCCAGC	7016
DB	193	GATCGAGGTACATCAAGACAAAACCCCTGCCACAGCTCAGTCATCAAGTGTGGTCCAGC	252
QY	7017	AAAAGCCACGACAGACTGCTCAGCCCTTCAGCTCGGCCCGCCAGCCCAACCCAGCCCA	7076
DB	253	AAAAGCCACGACAGACTGCTCAGCCCTTCAGCTCGGCCCGCCAGCCCAACCCAGCCCA	312
QY	7077	GTCCCCAGCTCAGCCTGAAGTTTCAGACTCAGCCTGAAGTTTCAGACCCCAACAACCTGTTTC	7136
DB	313	GTCCCCAGCTCAGCCTGAAGTTTCAGACTCAGCCTGAAGTTTCAGACCCCAACAACCTGTTTC	372
QY	7137	ATCCCATGTCCCTTCTGAAGACAAACCCACGACAGCTCATATCCAAAGCTTGC	7196
DB	373	ATCCCATGTCCCTTCTGAAGACAAACCCACGACAGCTCATATCCAAAGCTTGC	432
QY	7197	AGCACGCTCAGCCTCAAAAGTAAATGTCGAAGGACAGCTCCTGTTGCTGCCAAGTCC	7256
DB	433	AGCACGCTCAGCCTCAAAAGTAAATGTCGAAGGACAGCTCCTGTTGCTGCCAAGTCC	492
QY	7257	ATCACAGACTCGAATAGCTCCATCAACTCCATCCCACTGCTCCTGGACAAACATCCCA	7316
DB	493	ATCACAGACTCGAATAGCTCCATCAACTCCATCCCACTGCTCCTGGACAAACATCCCA	552
QY	7317	GGTTCAGACTACAACTCACAACCGATTCCAATTAACCAACATACATCTCTTCAGATACC	7376
DB	553	GGTTCAGACTACAACTCACAACCGATTCCAATTAACCAACATACATCTCTTCAGATACC	612
QY	7377	TTCCCAAGGCCAGCCAGTCACAACCCAGGTACAGTCTTCAACTCAAACTCTTTCATC	7436
DB	613	TTCCCAAGGCCAGCCAGTCACAACCCAGGTACAGTCTTCAACTCAAACTCTTTCATC	640
QY	7437	AGGACAAACTTTAAATCAAGTTAGTGTTCATCCCATCCCTCAGCTACAAATACA	7496
DB	641	AGGACAAACTTTAAATCAAGTTAGTGTTCATCCCATCCCTCAGCTACAAATACA	640
QY	7497	GCAGCCACGCCCAAGTCATTTGCTGTGCTCAGCTGCAACAACAGTCCAGGTTCTCTC	7556
DB	641	GCAGCCACGCCCAAGTCATTTGCTGTGCTCAGCTGCAACAACAGTCCAGGTTCTCTC	640
QY	7557	TCAGATCCAGTCACAGGTTTGCTCAGATACAGGCTCACAAGTGGTGTGCCACGCA	7616
DB	641	TCAGATCCAGTCACAGGTTTGCTCAGATACAGGCTCACAAGTGGTGTGCCACGCA	640
QY	7617	AATCAAACTCCAGTTACCTATCCAAATTCAGAAAGCAGTGTGTGCAGACTCACCAGAT	7676
DB	641	AATCAAACTCCAGTTACCTATCCAAATTCAGAAAGCAGTGTGTGCAGACTCACCAGAT	640
QY	7677	TCAGATGTGGTTACAGTGCAGGAGCCAGTGTGCAAGAGCAGTTGCAAAAGGTTACGCA	7736
DB	641	TCAGATGTGGTTACAGTGCAGGAGCCAGTGTGCAAGAGCAGTTGCAAAAGGTTACGCA	640
QY	7737	ACTCAGGGATCAGCAGCAAAAGAAAGAACAGACAGATACAAATTAACGTGAACACACC	7796
DB	641	ACTCAGGGATCAGCAGCAAAAGAAAGAACAGACAGATACAAATTAACGTGAACACACC	640
QY	7797	CTCCAAGCTTCTAATCAAAAGTTGAAATCATTCAGAAACAGGTGGTGTGATGAAGCATATGC	7856
DB	641	CTCCAAGCTTCTAATCAAAAGTTGAAATCATTCAGAAACAGGTGGTGTGATGAAGCATATGC	663

Db 1744 TGCACCATGGGAGAGATACAAAGACGATATATGAAAGCTGACGGAATTTGTGC 1803

QY 8823 AGATATGACCAAAATTTTGTATACTGTCGTACTACAAATCAAGTACTCCCAATTTA 8882

Db 1804 AGATATGACCAAAATTTTGTATACTGTCGTACTACAAATCAAGTACTCCCAATTTA 1863

QY 8883 CCAGTGTGAGAAGTCTCGAATCATCTTGTACAGAAATTCARAGGCTTCAAAGCTAG 8942

Db 1864 CCAGTGTGAGAAGTCTCGAATCATCTTGTACAGAAATTCARAGGCTTCAAAGCTAG 1923

QY 8943 CAGTCTCTAFAACAACTGAGTCTACAGTCTTAAAGTTCAGCGTGTAACTAA 9002

Db 1924 CAGTCTCTAFAACAACTGAGTCTACAGTCTTAAAGTTCAGCGTGTAACTAA 1983

QY 9003 CATAAACACAGCAAGAACTGTTGTCGTAAGTCTTAAATTAAGAGCCAGATGTTT 9062

Db 1984 CATAAACACAGCAAGAACTGTTGTCGTAAGTCTTAAATTAAGAGCCAGATGTTT 2043

QY 9063 TTAGTCAGGCTATCCTGACAGACTTGACCTAACTTCGTTTTATTGGTCAFAACAGTC 9122

Db 2044 TTAGTCAGGCTATCCTGACAGACTTGACCTAACTTCGTTTTATTGGTCAFAACAGTC 2103

QY 9123 CAATTATATCTTGGCCAAATTTTGTCCACGGACAAAGAAAGCAAGTCAACGACACC 9182

Db 2104 CAATTATATCTTGGCCAAATTTTGTCCACGGACAAAGAAAGCAAGTCAACGACACC 2163

QY 9183 ATTAATCTGACAGATCAGATGTTTGTGTCGTAAGTCTTAAATTAAGAGCCAGATGTTT 9242

Db 2164 ATTAATCTGACAGATCAGATGTTTGTGTCGTAAGTCTTAAATTAAGAGCCAGATGTTT 2223

QY 9243 TTGAAAAAAGAAAGAAAGAAAGCAAGAAAGAAAGTATGCGGTCAAGTCAACTCC 9302

Db 2224 TTGAAAAAAGAAAGAAAGAAAGCAAGAAAGAAAGTATGCGGTCAAGTCAACTCC 2283

QY 9303 ATGGAAATGCCAGCTGCTCTTCAGTGAAGAGCTGTTTAGAGTCTCACAGAAACTT 9362

Db 2284 ATGGAAATGCCAGCTGCTCTTCAGTGAAGAGCTGTTTAGAGTCTCACAGAAACTT 2343

QY 9363 TTGACTGTAATTAATTTGTCGAAAAAGAGAGCTTTTATTGTCGCGCTCATTTGTC 9422

Db 2344 TTGACTGTAATTAATTTGTCGAAAAAGAGAGCTTTTATTGTCGCGCTCATTTGTC 2403

QY 9423 AGCTAAGTATTTTCTTATAAAATCCAGCCCGGTTACATAATCATCTGATCTAT 9482

Db 2404 AGCTAAGTATTTTCTTATAAAATCCAGCCCGGTTACATAATCATCTGATCTAT 2463

QY 9483 CATGATTCCTGAGTAAAGTACAGACGACCTCTAGATGCTCTTTCTTCTATGAAG 9542

Db 2464 CATGATTCCTGAGTAAAGTACAGACGACCTCTAGATGCTCTTTCTTCTATGAAG 2523

QY 9543 GAGCTGCTATCTACACATGTGCACACACACAACTGGGAATCAACAAATGAGTTATTGT 9602

Db 2524 GAGCTGCTATCTACACATGTGCACACACACAACTGGGAATCAACAAATGAGTTATTGT 2583

QY 9603 TCATGGTATGATAAATTAAGTGTGATAAAGTGGCTTAAGTGGCTTGGCTACAG 9662

Db 2584 TCATGGTATGATAAATTAAGTGTGATAAAGTGGCTTAAGTGGCTTGGCTACAG 2642

QY 9663 ACTCTGTTGCTTGAATATACAGTACAAATTTGTCATCTCTGACCCAGGCTTAAAGTG 9722

Db 2643 ACTCTGTTGCTTGAATATACAGTACAAATTTGTCATCTCTGACCCAGGCTTAAAGTG 2702

QY 9723 AGTAAATCTATTTGAAGGTATCTTGTGTAACATTTGTCAGATCTTAATTTTTTCT 9782

Db 2703 AGTAAATCTATTTGAAGGTATCTTGTGTAACATTTGTCAGATCTTAATTTTTTCT 2762

QY 9783 TTTGTATTAATTAATTAAGTGTGATAAAGTGGATTAATTAATTAATTTTCT 9842

Db 2763 TTTGTATTAATTAATTAAGTGTGATAAAGTGGATTAATTAATTAATTTTCT 2822

QY 9843 CCCACAAAAA 9854

Db 2823 CCCACAGACAGA 2834

RESULT 9
BD157797 2057 bp DNA linear PAT 17-JAN-2003
LOCUS Primer for synthesizing full-length cDNA and use thereof.
DEFINITION
ACCESSION BD157797
VERSION BD157797.1 GI:27863555
KEYWORDS JP 2002191363-A/12640.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2057)
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 12640 09-JUL-2002
COMMENT
OS Homo sapiens (human)
PN JP 2002191363-A/12640
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
PI SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
(387)..(2057).
FT CDS Location/Qualifiers
source
1..2057
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 666 a 566 c 435 g 390 t
ORIGIN
Query Match 17.4%; Score 1719.4; DB 6; Length 2057;
Best Local Similarity 99.7%; Pred. No. 1.8e-265;
Matches 1723; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 5681 TAATTAAGAGGAGAGATGTTGGTCTTATGGCATTCGATTTGATTTGATTCAGGAAA 5740
Db 1 TAATTAAGAGGAGAGATGTTGGTCTTATGGCATTCGATTCGATTCGATTCAGGAAA 60
QY 5741 TCATTTGTCCTCATGAGTTCAGAAACACCAACCAAGAGCCCTACACCTCAGAGAAAG 5800
Db 61 TCATTTGTCCTCATGAGTTCAGAAACACCAACCAAGAGCCCTACACCTCAGAGAAAG 120
QY 5801 GCCTTCGATCAAGTGCAGTCCGCGCCAAAGAGACCAAGAGCCCAAGCAAACTGGCCCTG 5860
Db 121 GCCTTCGATCAAGTGCAGTCCGCGCCAAAGAGACCAAGAGCCCAAGCAAACTGGCCCTG 180
QY 5861 TTATTTGTAACCTGGGTAGCAGAGAACTGGAATTTGGGAGATCAGGCAATTTG 5920
Db 181 TTATTTGTAACCTGGGTAGCAGAGAACTGGAATTTGGGAGATCAGGCAATTTG 240
QY 5921 CTGAGAGTGGAGAAAAAGGACCAAGCAGTTGAGCAACAGGCTTAAGAAACGACTGG 5980
Db 241 CTGAGAGTGGAGAAAAAGGACCAAGCAGTTGAGCAACAGGCTTAAGAAACGACTGG 300
QY 5981 AGCAGAGAAAGCCGACAGTGAATTCGAACTTCCACTTCTCCCAACAGCAGTACACCA 6040
Db 301 AGCAGAGAAAGCCGACAGTGAATTCGAACTTCCACTTCTCCCAACAGCAGTACACCA 360
QY 6041 GCACCTCTCTCCAGCACAGAAAGTTATGTTGGCCCCCATAGTGGCTCAGTTTACAACTG 6100

361	DB	GCACCATCTCTCCAGCAGCAGAAGGTTAATGGTGGCCCCCAATAGTGGCTCAGTTTACAAC	420
6101	QY	GAACCAAAATGGTACTAACTACTAAAGTTGGATCTCCAGCTACAGTAACATTTCCAA	6160
421	DB	GAACCAAAATGGTACTAACTACTAAAGTTGGATCTCCAGCTACAGTAACATTTCCAA	480
6161	QY	ACAAGAACTTTTCATCAAAACCTTTTGCTACATGGGTTTAAGCAAGGCCAGTCAAA	6220
481	DB	ACAAGAGCTTTTCATCAAAACCTTTTGCTACATGGGTTTAAGCAAGGCCAGTCAAA	540
6221	QY	TTGTTCAAGTACAGCAGAAAAGTCCTGGGTATCATTCATCAAGTACAGGTACCAGT	6280
541	DB	TTGTTCAAGTACAGCAGAAAAGTCCTGGGTATCATTCATCAAGTACAGGTACCAGT	600
6281	QY	AAACCTTTTACTTTCATTCAGGCCAGGACGCAACAGTCAACAATTAGGCCCAAT	6340
601	DB	AAACCTTTTACTTTCATTCAGGCCAGGACGCAACAGTCAACAATTAGGCCCAAT	660
6341	QY	GCTCTGGAGAACCAACAAGAAATTCACAAGTAATCACAGGGCTTCAGATTCGCC	6400
661	DB	GCTCTGGAGAACCAACAAGAAATTCACAAGTAATCACAGGGCTTCAGATTCGCC	720
6401	QY	TGACCGTGATTAGAACACCACTCCACAGTCAACACTAGGAAGGCAATTTATTC	6460
721	DB	TGACCGTGATTAGAACACCACTCCACAGTCAACACTAGGAAGGCAATTTATTC	780
6461	QY	CTGTGTGTTGACAGCCAGGTGCTCCTCAGCAAGTGTAGTACTCAAACTATCAG	6520
781	DB	CTGTGTGTTGACAGCCAGGTGCTCCTCAGCAAGTGTAGTACTCAAACTATCAG	840
6521	QY	CTGTCTCCACTGCACTTCGCGCCCTTAACACAGGTTTCCTCAACACCTTGG	6580
841	DB	CTGTCTCCACTGCACTTCGCGCCCTTAACACAGGTTTCCTCAACACCTTGG	900
6581	QY	TAACCTTCAGCAAGTCCACTTCAAAATACAGTCTTACAGCCTCACACCCCT	6640
901	DB	TAACCTTCAGCAAGTCCACTTCAAAATACAGTCTTACAGCCTCACACCCCT	960
6641	QY	AACAAAGCACAAGTGAAGCTCACCATGGCTCAACTTACTCAGTTTAAACAG	6700
961	DB	AACAAAGCACAAGTGAAGCTCACCATGGCTCAACTTACTCAGTTTAAACAG	1020
6701	QY	GCAATCAAGGTTTGACAGTAGTAATTCAGGACAAGGTCAAACTACTTGG	6760
1021	DB	GCAATCAAGGTTTGACAGTAGTAATTCAGGACAAGGTCAAACTACTTGG	1080
6761	QY	TGATACCTCAAGGGGTGACTGTACTCCAGGCCAGGCGACGAGCTAATG	6820
1081	DB	TGATACCTCAAGGGGTGACTGTACTCCAGGCCAGGCGACGAGCTAATG	1140
6821	QY	TGCCAAATGGTACTGTTTCAGCGATTCTCTTTTACCCTATTGGCAACAAC	6880
1141	DB	TGCCAAATGGTACTGTTTCAGCGATTCTCTTTTACCCTATTGGCAACAAC	1200
6881	QY	CCAGCACCCACCAACCACTGTTTCCACGACGACGAGGTACAGGTGAACA	6940
1201	DB	CCAGCACCCACCAACCACTGTTTCCACGACGACGAGGTACAGGTGAACA	1260
6941	QY	GTAACCTGTACCCCGAGATCGAGGTACATCAAGCAAAAACCTTCCAC	7000
1261	DB	GTAACCTGTACCCCGAGATCGAGGTACATCAAGCAAAAACCTTCCAC	1320
7001	QY	CAAGTGTGGTCCAGCAAAAACCCAGCCACAGACTGCTCAGCCTTAC	7060
1321	DB	CAAGTGTGGTCCAGCAAAAACCCAGCCACAGACTGCTCAGCCTTAC	1380
7061	QY	CCCAAAACCCAGCCCGAGTCCCGCAGCTCAGCCTGAAGTTTCAGCT	7120
1381	DB	CCCAAAACCCAGCCCGAGTCCCGCAGCTCAGCCTGAAGTTTCAGCT	1440
7121	QY	CCCAAAACCACTGTTTCAATCCCATGTCCCTTCTGAAGCAACAAC	7180
1441	DB	CCCAAAACCACTGTTTCAATCCCATGTCCCTTCTGAAGCAACAAC	1500

Qy	7181	CCAAAGCCCCAAGTTGCAAGCAGACTCTCAGCCTCAAGTAATGTGCCAAGGACAGTCTCCTG	7240
Db	1501	CCAAAGCCCCAAGTTGCAAGCAGACTCTCAGCCTCAAGTAATGTGCCAAGGACAGTCTCCTG	1560
Qy	7241	TTCGTGTCAAAGTCCATCAGACACTCGAATAGTCCATCAACTCCATCCCAACTGTCTC	7300
Db	1561	TTCGTGTCAAAGTCCATCAGACACTCGAATAGTCCATCAACTCCATCCCAACTGTCTC	1620
Qy	7301	CTGGAACAACATCCCAGGTTTCAGACTACAACCTCACAACCGATTCCAATTTCAACCCACATA	7360
Db	1621	CTGGAACAACATCCCAGGTTTCAGACTACAACCTCACAACCGATTCCAATTTCAACCCACATA	1680
Qy	7361	CATCTCTTCAGATACCTTCCCAAGGCGACCAAGTCACAACCCCCAGGT	7409
Db	1681	CATCTCTTCAGATACCTTCCCAAGGCGACCAAGTCACAACCCCCAGGT	1729

RESULT 10

AK022570
LOCUS

FOCUS DEFINITION

ACCESSION

VERSION

KEYWORDS
SOURCE

ORGANISMS

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE
AUTHORS

TITLE	DATE	BY	NO.	PRICE	REMARKS
...

JOURNAL

FEATURES

source

BASE COIN

BASE COUNT
ORIGIN

Quercus M.

Query M
Best Loc

Matches

Qy

10

DD

5741	Qy	TCATTTGTCCTCCATTTGGAGTTTCCAGAAACACCAAAAGAAAGCCCTACACCTCAGAGGAAAG	5800
61	Db	TCATTTGTCCTCCATTTGGAGTTTCCAGAAACACCAAAAGAAAGCCCTACACCTCAGAGGAAAG	120
5801	Qy	GCCTTCGATCAAGTGCACCTCGGCCCAAGAGACCCAGAAAGCCCAACCAACCTGCGCCCTG	5860
121	Db	GCCTTCGATCAAGTGCACCTCGGCCCAAGAGACCCAGAAAGCCCAACCAACCTGCGCCCTG	180
5861	Qy	TTATTTATTGAAACCTGGGTAGCAGAGAAGAACTTGGAAATTTGTGGAGATCAGGGCATTTG	5920
181	Db	TTATTTATTGAAACCTGGGTAGCAGAGAAGAACTTGGAAATTTGTGGAGATCAGGGCATTTG	240
5921	Qy	CTGAGAGTGGAGAAAGAAAGGACAGAGCTTTGAGCAACAGGCTTAAGAAACGACTGG	5980
241	Db	CTGAGAGTGGAGAAAGAAAGGACAGAGCTTTGAGCAACAGGCTTAAGAAACGACTGG	300
5981	Qy	AGCAGCAGAGCCGACAGTCAATTTGCCAATTTCCACTACTTCCCCCAACAGCAGTACACCA	6040
301	Db	AGCAGCAGAGCCGACAGTCAATTTGCCAATTTCCACTACTTCCCCCAACAGCAGTACACCA	3600
6041	Qy	GCACCACTCTCCAGCACAGAAAGTTATGTGTGGCCCCCATTAAGTGGCTCAGTTTACAACCTG	6100
361	Db	GCACCACTCTCCAGCACAGAGGTTATGTGTGGCCCCCATTAAGTGGCTCAGTTTACAACCTG	420
6101	Qy	GAACCAAAATGGTACTAACTACTAAAGTTGGATTCCTCAGCTACAGTAACATTTCCAACAAA	6160
421	Db	GAACCAAAATGGTACTAACTACTAAAGTTGGATTCCTCAGCTACAGTAACATTTCCAACAAA	480
6161	Qy	ACAAGAACTTTTCATCAAACTTTTGCTACATGGGTTTAGCAAGGCCAGTCAAAATTCAGGCG	6220
481	Db	ACAAGAACTTTTCATCAAACTTTTGCTACATGGGTTTAGCAAGGCCAGTCAAAATTCAGGCG	540
6221	Qy	TTGTTCAGGTACAGCAGAAAAGTCTCGGGTATCATTTCCATCAAGTACAGGTACCAGTCAAGC	6280
541	Db	TTGTTCAGGTACAGCAGAAAAGTCTCGGGTATCATTTCCATCAAGTACAGGTACCAGTCAAGC	600
6281	Qy	AAACCTTTACTTCAATTCAGCCGAGGACAGCAACAGTCAAAATTTAGGCCCAAATACCTTCAG	6340
601	Db	AAACCTTTACTTCAATTCAGCCGAGGACAGCAACAGTCAAAATTTAGGCCCAAATACCTTCAG	660
6341	Qy	GCCTCGAGAACCCACAGCAAAATTCACAAGTAACTACAGGCCCTCAGATTCGCGCCTGGTA	6400
661	Db	GCCTCGAGAACCCACAGCAAAATTCACAAGTAACTACAGGCCCTCAGATTCGCGCCTGGTA	720
6401	Qy	TGACCGTGATTAGAACACCACCTCCAACAGTCAACACTTAGGAAAGGCCAAATTTATTCGAACAC	6460
721	Db	TGACCGTGATTAGAACACCACCTCCAACAGTCAACACTTAGGAAAGGCCAAATTTATTCGAACAC	780
6461	Qy	CTGTGATGGTACAGCCAGGTGCTCCTCAGCAAGTGATGACTCAAACTATCAGGGGGCAGC	6520
781	Db	CTGTGATGGTACAGCCAGGTGCTCCTCAGCAAGTGATGACTCAAACTATCAGGGGGCAGC	840
6521	Qy	CTGTCTCCACTCGAGTCTCGGCCCTTAACAGGTTTTCCTCAACACCTGGCAGAAAAGCT	6580
841	Db	CTGTCTCCACTCGAGTCTCGGCCCTTAACAGGTTTTCCTCAACACCTGGCAGAAAAGCT	900
6581	Qy	TAACCTTCAGCAACGCTCCACTTTCAAATATACAGTCTTCAGGCTTCACAACCCCTCGCCCC	6640
901	Db	TAACCTTCAGCAACGCTCCGCTTTCAAATATACAGTCTTCAGGCTTCACAACCCCTCGCCCC	960
6641	Qy	AACAAGGACAAGTGAAGCTCACCATGGCTCAAACTTACTCAGTTAAACAGGGCCACGGTG	1020
961	Db	AACAAGGACAAGTGAAGCTCACCATGGCTCAAACTTACTCAGTTAAACAGGGCCACGGTG	1080
6701	Qy	GCAATCAAGGTTTGACAGTAGTAATTTCAAGGACAAAGTCAAACTTACTGGACAGTTGCAGT	6760
1021	Db	GCAATCAAGGTTTGACAGTAGTAATTTCAAGGACAAAGTCAAACTTACTGGACAGTTGCAGT	1080
6761	Qy	TGATACCTCAAGGGGTGACGTACTCCAGGCCCCAGGCCAGCAGTCAATTCAGCTGCAAA	6820
1081	Db	TGATACCTCAAGGGGTGACGTACTCCAGGCCCCAGGCCAGCAGTCAATTCAGCTGCAAA	1140
6821	Qy	TGCCAAATGGTACTGTTTCAGCGAATTCCTCTTTACCCCAATTTGGCAACAAACAGCCACACAG	6880

[illegible]

RESULT 11	BC032289	2549 bp	mRNA	linear	ROD 22-APR-2003
LOCUS	BC032289				
DEFINITION	Mus musculus fetal Alzhelmer antigen, mRNA (cDNA clone IMAGE:5368380), partial cds.				
ACCESSION	BC032289				
VERSION	BC032289.1	GI:21619020			
KEYWORDS					
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 2549)				
	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schaefer, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullighy, S.J., Bosak, S.A., McEwan, P.J., McKernan, R.K., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, N.D., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madañ, A., Rodrigues, S., Sanchez, A., Whitting, M., Madañ, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.				
	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences				
	TITLE				

Db 972 TTGTATCAGAGATCATTTTGGCCCATTTGGAGTTCAGAGCACCACCAAAAGAACACCCAC 1031

Qy 5787 ACCTCAGAGAAAGGCTTCGATCAAGTGCACACTTGGCCCAAGAGACCAAGAACGCCCAA 5846

Db 1032 ACCTCAGAGAAAGGCTTCGATCAAGTGCACACTTGGCCCAAGAGACCAAGAACGCCCAA 1091

Qy 5847 GCAAACTGGCCCTGTTATTATTGAAACCTGGGTAGCAGAGAGAACTGGAATTTGGGA 5906

Db 1092 GCAGACGGGCCCTGTTATCATTCAGAGCTGGGTAGCAGAGAGAGCTGGAGCTCTGGGA 1151

Qy 5907 GATCAGGGCATTTGCTGAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5966

Db 1152 GATCAGGGCTTTTGGCTGAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1211

Qy 5967 TAAGAACGACTGGAGCAGAGAGCGAGAGTATTGCAACTTTCACACTTACTTCCCAAC 6026

Db 1212 CAGAAACGTTTGGAGCAG 1271

Qy 6027 AAGCAGTACAAACAGCAGCAGTCTCTCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 6086

Db 1272 AAACAACAGCAGCAGCAGTCTCTCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1331

Qy 6087 CTCAGTTACAACTGGACCAAAATGTACTAACTACTAAAGTTGGATCTCCAGCTACAGT 6146

Db 1332 CTCGGTACCCCTGGACCAAAAGATGGTGGCCACAAAGTGGCTGGCTGCCACAGT 1391

Qy 6147 AACATTCCAAACAAAGAACTTTCATCAAACTTTGTACATGTTGGTTAAGCAAGGCCA 6206

Db 1392 CACATTCAGCAAAACAGAACTTTCATCAAACTTTGCCACGTGGTGAAGCAAGGCCA 1451

Qy 6207 GTCAAATTCAGGCGTTGTTCAAGTACAGCAGAGAGAGTCTGGGTATCATTCCTCAAGTAC 6266

Db 1452 GTCCAAATTCAGGATGTTCCAGGTACAGCAGAGAGAGTCTGGGATCATTCATCGACTAC 1511

Qy 6267 AGTACAGTACAGAACCTTTACTTCATTCAGCCAGCCAGCAGCAGCAGCAGCAGCAGT 6326

Db 1512 AGGTCAAGCCAGCAAACTTTACTTCATTCAGCCAGCCAGCAGCAGCAGCAGCAGT 1571

Qy 6327 GCCCAATACCTCAGGCTCTGGAGGAAACCAAGCAATTCACAACTATCAAGTACAGGGCTCA 6386

Db 1572 GCCCAATACCTCAGGCTCTGCAGGAACTACTACCAATACAGTATCAAGTACAGGGCTCA 1631

Qy 6387 GATTGCGGCTGATGACCGTGTATGACACACACTCCACAGTCAACACTAGAGAGGC 6446

Db 1632 GATCCGCGCTTGGATGACAGTGTATGAGCGCCACTCCAGCAGCCAGCAGTGGGAAGGC 1691

Qy 6447 AATTATTGCAACACTCTGTATGTTACAGCAGGCTGCTCCTCAGCAAGTGTACTCAAT 6506

Db 1692 GATCATTCGAACACACAGTGTGTGTCAGCCAGGCACACCCACCAAGTTGTGACTCAAT 1751

Qy 6507 CATCAGGGGAGCGCTGTCTCCACTGCAGTCTCCGCCCTTAACACGCTTTCTCACAACC 6566

Db 1752 CATCCGAGGAGCGCTTTTCCACTGCCATCTGTCTCCAAGTACAGTCTCCTCAGCAAC 1811

Qy 6567 TGGCAGAAAGCTTAATTCAGAGAGTCCACTTCAATATACAGTCTTCAGCTCACA 6626

Db 1812 TGTACAAAGGGCTTAATTCAGAGAGCGGCTGGGCCCTTCAGCGCTTCAGCCCAACA 1871

Qy 6627 ACCCCCTCGCCCAACAGGAGCAAGTGAAGCTCACCATGGCTCAACTTACTCAGTTAAC 6686

Db 1872 CAGCCCTCGCTCAGAGAGGAGGAGTGAAGCTCACCATGGCTCAGTCACTCAGCTCAC 1931

Qy 6687 ACAGGGCCAGGTTGGCAATCAAGTGTTCAGAGTAGTAAATTCAGAGCAAGGTCAAACTAC 6746

Db 1932 GCAGGGCCATGGGGCAACCAAGGTTTAAACAGTAGTGAATTCAGAGCAAGGGCAGACTAC 1991

Qy 6747 TGGACAGTTGCAAGTTGATACCTCAAGGGGTGACTGTACTCCAGGCCAGCCAGCAGCT 6806

Db 1992 TGGACAGTTGCAAGTTGATACCTCAGGGGATGACAGTCTTCCAGGGCCAGGGCAGCAGCT 2051

Qy 6807 AATCAAGCTGCAATGCAAAATGGTACTGTTACAGCGATTCTCTTTTACCCCAATTTGGCAAC 6866

Db 2052 CATGCAAGCTGCAATGCGCAATGGCAGCTTTCAGCGTTTCTCTTTACCCCACTATCAAC 2111

Qy 6867 AACAGCCACCAGCAGCCAG-----CACCAACCACACCACCACTGTTTCCACGACAGC 6914

Db 2112 ATCAGCCACGGCGCCAGCAGCAGCAACAGCAGCAGCAGCAGCACTACTTAATGCTACAGC 2171

Qy 6915 ACAGGTACAGGTGAACAAGCAGAGTAACTGTCCACCCAGATGCAGGTACATCAAGA 6974

Db 2172 AGCAGGCTCAGGGGAACAAAACAGAGTAAATATTTACCCAGACCCAGGTACAAACAGC 2231

Qy 6975 CAAAACCCCTGCCACAGCTCAGTCATCAAGTGTGGGTCCAGCAAAAGCCAGCCAGCAGAC 7034

Db 2232 CACAACCCCTGGCAGCAACTCAGTCATCAAGTGTGAGTCTCGCAGAGCCAGCCAGCC 2291

Qy 7035 TGCTCAGCCCTTCAGTCGGCCGCCAGCCCAACCCAGCCAGTCCCCAGCTCAGCCTGA 7094

Db 2292 TGCTCAGCCTGCAGCAACAACCCAGCCAGCCAGCCAGCC----- 2331

Qy 7095 AGTTCAGACTCAGCCTGAAGTTCAGACCCCAACAACTGTTTCATCCCTCCCTCTCTGA 7154

Db 2332 ----CAGCTCAGCCTGAAGTTCAGACCCAGCCAGCCAGCAGTGTCTGTCCTATGTCCTCTGA 2387

Qy 7155 AGCAACAACCCACCC 7168

Db 2388 AACACAGCCCTCCC 2401

RESULT 12
BC046973 2769 bp mRNA linear ROD 16-APR-2003
LOCUS
DEFINITION
Mus musculus hypothetical protein LOC268496, mRNA (cdna clone
MG:54978 IMAGE:6487444), complete cds.

ACCESSION
BC046973
VERSION
BC046973.1 GI:28422395
KEYWORDS
MGC.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

REFERENCE
1 (bases 1 to 2769)
AUTHORS
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Usdin,T.B., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Ussin,T.B., Loquellano,N.A., Peters,G.J.,
Carninci,P., Prange,C., Raha,S., Bosak,S.A., McEwan,P.J.,
Abramson,R.D., Mullahy,S.J., Gunaratne,P.H., Richards,S.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmitz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalios,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL
MEDLINE
PUBMED
22388257
12477932

REFERENCE
2 (bases 1 to 2769)
AUTHORS
Strausberg,R.
TITLE
Direct Submission
JOURNAL

Submitted (14-FEB-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: qcapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.

REMARK
COMMENT

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nigr.nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, R.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
 Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 100 Row: m Column: 22
 This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction, Similarity but not identity to protein.

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10

```

1. .2769
/organism="Mus musculus"
/mol_type="mrna"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="MGC:54978 IMAGE:6487444"
/tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month
old mouse. Taken by biopsy."
/clone_lib="NCI-CGAP_Mam2"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. .2769
/gene="LOC268496"
/db_xref="LocusID:268496"
163. .2577
/codon_start=1

```

/product="LOC268496_protein"
 /protein_id="AAH46973.1"
 /db_xref="GI:28422396"
 /db_xref="LocusId:268496"
 /translation="MYSEDEDEEDEEDDAEETDQSDEREEDDEDDSDYPEEMED
 DDHDSYCTESSPRSHSTYSSTPGRRKPVHRPSPILLEKDIPLLEFPKSSDLMVP
 NEHIMNVIAIYEVVRNGVNLRLSPCEFCFCAALVSGQCTLMAEHWKALLKAVILRE
 EDSNTTGPADLKDSYNTLTFIDGNTWPELVRYCESDKTHHVLPTQEAADYPIG
 PVENKILQLFVDDQFLTNIAAREELMSEGVIOYDDHRCVCHKLGLDLCCTCSAYIH
 LECVKPLPVEDEWOCEVCYVAHKVPYGTDCVAYEOKNKPVYRHEPIGYDRSRRIYH
 FLNRRLLI EEDTDNEKKVWYISTKVLAELDCLDQGYWEALCRVLIEDIRMEQGN
 HMDVTDLTNKARGSNKFLAANEBILDSLRKRGEDIDCQSPEDPKDKHEGNN
 SKDKASKEEAEDPSADKDQSKLEPEEGHGKPRDEDSNRTCDTSNTSATTAASSOPN
 LFTCSSELTSSQSDSKAADDPDIERGSPSTPVSHREDIGDFRLKSVYVNSQGEVSRLL
 GKGTSSTAIRILPLRNPEKLSOLKSOQVAAAHAANKLFEKGEKVLVYNSQGEVSRLL
 STYKEVVMGGINNYFKLQGEGRYVIHQISTNSFALNKHQREDDHNRHLIAHQAQMC
 LTPAGEFKWNGSVHGSKSTITSLRTITIQLESNI PPSFLHPNWSHRRNWKIYAVQMC
 SKPREFALAILALECAVKPVVMLPIWRESLGHTRLHRMTSIEREEKEKVKKKKKKKGR
 PFFFFF"
 862 a 609 c 753 g 545 t
 BASE COUNT
 ORIGIN

```
Query Match      17.3%; Score 1708.4; DB 10; Length 2769;
Best Local Similarity 80.0%; Pred. No. 1e-263;
Matches 2193; Conservative 0; Mismatches 336; Indels 213; Gaps 7;
```

96	CAGCACCAAGCCCCGGGGCCGGGGGGGGAGAGAGCGGGGGCGGCACGACGGGGGGCGG	155
QY		
6	CAGCGCAGCGCCCCGGGGCCGGGGGGCGAGAGCGGGGGCGGCAGACGGGGGGCGG	65
Db		
156	GGCGCGCGGGGCCACTGTCTCCGGACACCGCGGGCCCGAGGGCGCCGTCAACAAAGTGGT	215
QY		
66	G-----GGCGGCACCTTGCGCGGACACACCGCGGGCCCGCGCGCGTCAACAAAGTGGT	119
Db		

Qy	216	GTACGATGACCAACGAGACGAGCGGTGGAGGAAGAGGAGACATGCTCTCCGAG-----	270
Db	120	GTACGATGACCAACGAGAGTACACGACGACGAGGAGGAGATGCTCTCGGAGGAGGA	179
Qy	271	-GAGGAGGAGGAGGACGCGGACGCGGAGGAGACCCAGGATCTCTGAGACGACGAGGA	329
Db	180	CGAGGAGGAGGAGGATGGCGACGCGGAGGAGACCCAGGATCTCGGAGGACGAGGAGGA	239
Qy	330	GGATGAGATGGAAGAGGACGACGATGACTCCGATTTATCCGAGGAGATGGAAGACGACGA	389
Db	240	AGATGACATGGAAGAAGCAGCAGTACTCCGATTTATCCGAGGAGATGGAAGCAGCAGGA	299
Qy	390	CGAGCAGCCAGTTTACTGACGGAAGACGCTTCAGGAGCCATAGTACTACGACGACAC	449
Db	300	CGACGACGCTAGTTACTGTCACGGAAAGCAGCTTCAGGAGCCATAGTACATACGACGACAC	359
Qy	450	TCCAGGTAGCGAAACCAAGAGTACATCGGCCTCGTTCTCCATATATTGGAAGAAAGA	509
Db	360	TCCAGGTAGCGAAACCAAGAGTACATCGGCCTCGTTCTCCAAATATTAGAAGAAAGA	419
Qy	510	CATCCCCCCCTTTGAATTTCCCAAGTCTCTGAGGATTTAATGGTGCTTAATGAGCATAT	569
Db	420	CATCCCCACCCCTTGAATTTCCGAAGTCTCTGAGGATTTAATGGTGCTTAATGAGCATAT	479
Qy	570	AATGAATGTCAATGGCCATTTACGAGGTACTGCGGAACCTTTGGCACTGTTTTGAGATTATC	629
Db	480	AATGAATGTCAATGGCCATTTATGAGGTGGTGGCAACCTTTGGCAACCTTTAAAGTTGTC	539
Qy	630	TCCTTTTCGCTTTGAGGACATTTTGTGCAGCTCTGGTGAGCCAGAGCAGTGCACACTCAT	689
Db	540	CCCTTTTTCGCTTTGAGGACATTTCTGTGCGGCTCTGGTGAGCCAGAGCAGTGCACACTCAT	599
Qy	690	GGCAGAGATGCATGCTTGCTTTTGAAGCAGTTCTGCGTGAAGAGACACTTCCCAATAC	749
Db	600	GGCGGAGATGCAGTGGCGCTTCTGAAGACAGTTCTTCGAGAGAGGACACGCTCTACAC	659
Qy	750	TACCTTTGGACCTGCTGATCTGAAGAATAGCGTTAATTCACACTCTGTTTTCATAGATGG	809
Db	660	CACCTTTGGCCCTGCGGACTTGAAGACAGTGAACCTCCACCTGTATTTTCATCGACGG	719
Qy	810	GATGACGTGGCCAGAGTGTCTGGGTGTACTGTGAGAGTGATTAAGAGGTACCATCACGT	869
Db	720	GATGACGTGGCCAGAGTCTCTGAGGTFCTACTGTGAGAGTGACAAGAGGTACCACCATGT	779
Qy	870	TCTTCTCTTACCAAGAGCAGAGGACTACCCATATGGACCAGTAGAGAACAAAGATCAAAGT	929
Db	780	CCTTCTCTTACCAAGAGCAGAGGACTATCCCTACGGACCAAGTGGAGNACAAAATCAAAGT	839
Qy	930	TCTACAGTTTCTTAGTCGATCAGTTCTTCAACAAATATTTGCTCGAGAGGAATGTATGTC	989
Db	840	TCTGCAGTTTCTAGTCGATCAGTTCTTACGACCAATATTTGCTGCTGAGGAGCTGATGTC	899
Qy	990	TGAAGGGTGATACAGTATGATGACCATTGTAGGGTTTGTTCACAAACTTTGGGGATTTGCT	1049
Db	900	TGAAGGGGTAAATCCAGTATGACGACCAATCTGAGGGTCTGTCAAAAATTTGGGGATCTGCT	959
Qy	1050	TTGCTGTGACATGTTTCAGCAGTATACCAATTTTGAATGTGTGAAGCCACTCTTGAGGA	1109
Db	960	GTGCTGTGACAGCTGCTCGCTGTGTACCATTTAGATGTGTGAAGCCGCTCTTGNAGA	1019
Qy	1110	GGTGCCAGAGGACGAGTGGCAGTGTGAAGTCTGTGTAGCACACAAAGTGCTCGGTGTGAC	1169
Db	1020	AGTGCCGGAGGACGAGTGGCAGTGTGAAGTCTGTGTGGCACACAAAGGTGCTCGTGTGAC	1079
Qy	1170	TGACTGTGTTGCTGAAATCCAAAAAATAAACCATATATTTTCGACATGAACTTATTCGATA	1229
Db	1080	TGACTGTGTGCTGAAGTCCAAAAGATAAACCCCTACGTTTCGACAGCAACTATTGGGTA	1139
Qy	1230	TGATAGAAGTCGGAGGAATACTGGTTCTTGAACCGAAGACTCATATAGAACAAAGATAC	1289
Db	1140	TGACAGAACCGTAGGAAATAC'TGGTTTTTAAACCGAAGACTCATATAGAACAAAGACAC	1199
Qy	1290	AGAAATGAAATGAAAAGAAAAATTTGGTATTACAGCACAAAGGTCCAACTTCGCAAAAT	1349

Db	1200	AGACAATGAAAACGCAAGAAAGTTTGGTACTACAGCAAAAGGTCACGTTAGCAGAGCT	1259
Qy	1350	AATTGACTGCTAGACAAAGATTATTGGGAAGCAGAACTCTGCAAAATTTCTAGAGAAAT	1409
Db	1260	AATAGACTGGCTTAGACAAAGGATAC TGGGAAGCAGAACTGTGCAGATTTCTAGAGACAT	1319
Qy	1410	GCGTGAAGAAATCCACCACACATGACATAACTGAAGACCTGACCAATTAAGGCTCGGG	1469
Db	1320	TCGGAGGAGATGCAGCAGCATATGATGTAACGGGAAGACCTGACCAACAAGCTCGTGG	1379
Qy	1470	CAGTAACAATCCCTTTCTGGCGCAGCTTAATGAGAAATTTTGGAAATCCATTAAGACCAA	1529
Db	1380	CAGCAACAAGTCCTCTCTAGCGCAGCTAATGAGGAAATTTTGGACTCCTTTAAGAAATCAA	1439
Qy	1530	---AAAGGGAGACATTGTAATGTTTAAAGCCACAGAGAAACAGAAAAAGACAAAGAATGA	1586
Db	1440	ACGAGGGAAGATATTGACTGTGATCAGAGCCCGAAGACCCAGAAAGACNAGCATGA	1499
Qy	1587	GACTGAGAATGA---CTCTAAAGATGCTGAGAAAAACAGAGAAGAAATTTGAAGACCAAGTC	1643
Db	1500	GGCGAGAACAAACAGCTCCAAAGATGCTGAGAAAGCAGGAGGAGCGGAGACCTTTC	1559
Qy	1644	CCTTGAAAAGACAGTGAACACAAAAACACAGATGATGACCTTGACCAAGGAAAAATCTGA	1703
Db	1560	CGCAGACAGGACGCTGACAGCAAAAGCCCTGGAGGAAGAGCCCGGCATGGAANAACCTGA	1619
Qy	1704	GGAGCCAAACAGAAGTTGGGGATAAAGGTAAGCTGCTGTGCAGCAAAATCTTTGGGACAAAC	1763
Db	1620	G-----	1620
Qy	1764	AACAAATGCAATTCAGAAAGACTAGTCCCTCTGAAGGAGGAGGCCCTGTGGGTGTCT	1823
Db	1621	-----	1620
Qy	1824	CTCAGAAACCCCGATAGCAGCAACATGGCAGAGAAGAAGGTGGCATCTGAGCTCCCCCA	1883
Db	1621	-----	1620
Qy	1884	GGATGTCCAGAGAAGAACTTAACAAGACATGTGAGAGCAGTAACACTAGTCTACCACATAC	1943
Db	1621	-----GAAGACTCTAACAGACATGTGACAGCAGTAAACACATAGTGTACCACATGC	1670
Qy	1944	CTCCATCCAGCCTAATCTGAAAAACAGTAACACAGCAGTGAACATAAATTTCTCCAGAG	2003
Db	1671	CTCAGCCAGCCTAACTGGAGACCTG--CAGCAGCAGCAGCTCACTTCTTCCAGAG	1727
Qy	2004	TGAATCTGTAAGCAGCTGATGATCCTGAAAATGGAGAAAGAAATCTCATACACCTGT	2063
Db	1728	TGACTCTGCTAAGCAGCGGATGACCTCGAGATTGGAGAAAGAGACTCCCATACACCTGT	1787
Qy	2064	CTCTATTTCAGGAAGATAGTGTAGTGATTTTCACATCGGAGAAGTCCACCGGGAGCTAAG	2123
Db	1788	CTCTGTTACGAAGAGA---TAGGTGATTTTCAGACTGGAGAGTCCANTGGAGAAGTTTC	1844
Qy	2124	TGAATCTCTGGAGCTGGAAAAGAGGACATCTGGCTCAACTTCGAATCATACAGATTGCG	2183
Db	1845	CGAGTCCCTCGAGCTGGGAAAAGAACATCTGGGTCCACTCGTATCATCACAGACTGAG	1904
Qy	2184	GAATCCAGATAGCAAACTTAGTCAGCTGAAGAGCCAGCAGGTGGCAGCCGCTGCACATGA	2243
Db	1905	GAACCCGGAGACAAATCTAGTCACTCAGAGTCAGAGTGGCAGGTGGCAGCTGCAGCCATGA	1964
Qy	2244	AGCAAAATAATTTATTTAAGGAGGCCAAGAGGTACTGGTAGTTAACTCTCAAGGAGAAAT	2303
Db	1965	AGCAAAATAATTTTAAAGGAGGGAAGAGGTGCTGGTGTAACTTCCCAAGGAGAGGT	2024
Qy	2304	TTACAGGTTGAGCACCAAAAAGGAGTGTATCATCAAAAGGAATATCAACAATATTTTAA	2363
Db	2025	TTACAGGCTGAGCACCAAAAAGGAGGTGGTCATGAAAGGGAACATCAACAACATATTTTAA	2084
Qy	2364	ATTGGGTCAAGAAGGAAGTATCCGCTCTACCACAATCAATCTCCACCAATTCATTTGC	2423

Db	2085	GTGGGGACAAGAGGGGAAGTATCGAGTCTACCAACAATCACTACGACCAATTCATTTCG	2144
Qy	2424	TTTGAATTAAGCACACAGACACAGAAAGACCATGATAAGAGAAGGCATCTTGACACATAAGTT	2483
Db	2145	TCGAAATAACACACAGCACAGAGAAGACCATGATAAGAGGAGGCATCTTGACACATAAAT	2204
Qy	2484	CTGTCTGACTCCAGCAGAGAGAGTTCAATGGAACGGTCTGTCCATGGGTCCAAAGTTCT	2543
Db	2205	CTGTCTCACTCCAGCTGGGAGTTCAATGGAACGGGTCTGTTCACCGGTGCCAAAGTCCCT	2264
Qy	2344	TACATATCTACTCTGAGACTGACTATCACCCAATTAGAAAACAACATCCCTTCATCCTT	2603
Db	2265	TACCATCTCAACCCCTGAGACTCACCATCAACACACTAGAGAGCAACATTCCTTCATCCTT	2324
Qy	2604	TTTTTCATCCCAACTGGCGCATCACATAGGCGCAAAATGGATCAAGCCAGTTTCAGATGTGTAG	2663
Db	2325	CCTTACCCCAATTGGGCTTCACACAGGGCTAAATTGGATCAAGCCCGTTTCAGATGTGTAG	2384
Qy	2664	CAAAACCCAGAGAAATTTGCATTGGCTTTAGCCATTTTGGAGTGTGCAGTTAAACCAGTTGT	2723
Db	2385	TAAACCCAGAGAGTTTCGCTGGCTTCGCAATTCCTGGAGTGTGCAGTTAAGCCAGTTGT	2444
Qy	2724	GATGCTACCAATATGGCGAGAAATTTTAGGACATACCAAGTTTACACGGGTGACATCAAT	2783
Db	2445	GATGCTGCCAATATGGCGGGAGTCTCTAGGACATACCCGGTTACATAGGATGACATCAAT	2504
Qy	2784	TGAAGAGAGAGAAAAGGAGAAAGTCAAAAAAAGAGAGAA	2825
Db	2505	TGAAGGGAGAGAAAAGAGAAAGTTAAAAAAGAAAAA	2546

RESULT 13	BC037661	2481 bp	linear	ROD 16-APR-2003
LOCUS	BC037661			
DEFINITION	Mus musculus hypothetical protein LOC268496, mRNA (cDNA clone IMAGE:5353895), partial cds.			
ACCESSION	BC037661			
VERSION	BC037661.1			
GI:	22902292			

SOURCE
Mus musculus (house mouse)

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 2481)

AUTHORS
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Wax, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raja, S., Sack, S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Faney, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shenvchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Maria, M.A.

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
JOURNAL

Generation and initial analyses of more than 27,000 human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 2481)
Strausberg, R.
Direct Submission
Submitted (12-SEP-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

JOURNAL

Submitted (08-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 223387)
 Rat Genome Sequencing Consortium.
 Direct Submission

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (08-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Oct 9, 2002 this sequence version replaced gi:21738143.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: GOIO
 Center clone name: CH230-303015

----- Summary Statistics

Assembly program: Phrap; version 0.990329
 Consensus quality: 191060 bases at least Q40
 Consensus quality: 195253 bases at least Q30
 Consensus quality: 198285 bases at least Q20
 Estimated insert size: 194499; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved

* 1 222161: contig of 222161 bp in length
 * 222162 222261: gap of unknown length
 * 222262 223387: contig of 1126 bp in length.

FEATURES

source

Location/Qualifiers

1. .223387
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-303015"

misc_feature

1. .1317
 /note="wgs_end_extension"

misc_feature

clone_end:T7
 complement(1752..2639)
 /note="clone_boundary"

misc_feature

clone_end:T7
 site:MboI
 end_sequence:RXAKI92TJ

misc_feature

4700..5756
 /note="wgs_contig"

misc_feature

29062..30112
 /note="wgs_contig"

misc_feature

43664..45232
 /note="wgs_contig"

misc_feature

101380..102634
 /note="wgs_contig"

misc_feature

195324..197239
 /note="wgs_contig"

misc_feature

complement(215803..216669)
 /note="clone_boundary"

misc_feature

end_sequence:RXAKI92TV
 220191..222161
 /note="wgs_end_extension"

BASE COUNT 52000 a 44167 c 46629 g 57824 t 22767 others
 ORIGIN

Query Match 14.7%; Score 1454.4; DB 2; Length 223387;
 Best Local Similarity 76.1%; Pred. No. 5.9e-223;
 Matches 1833; Conservative 0; Mismatches 566; Indels 9; Gaps 3;

QY	3210	AGATGTGAAGGAGCTCTTAGATCTGACAGTGTAAACCTCGAAGGAAGCAATGGA	3269
Db	46425	AGATGTAAAGAAGACTTTTAGATCTGTAAATGACAAACCTTCAAGGAAGCAATGGA	46484
QY	3270	AGTAGACGATGACATGAAACAGAGTCACATGTAATTTGTCAGGAGAGTTCTCAAGTGA	3329
Db	46485	GATCGATGACAGTGTGAAACAGAGTCACATATGACAGTGAAGAGAGTGCACAGGTGA	46544
QY	3330	TGTGGTCAATGTAGTGAGGGTTTTTCATCTAAGGACTAGTTACAAAAAGAAAAACAATC	3389
Db	46545	TGTGGTCAATGTAGTGAGGGTTTTTCATCTAAGGACTAGTTACAAAAAGAAAAACAATC	46604
QY	3390	ATCCAACTAGATGGACTTCTTGAAGGAGAAATTAACAGTTTACACTGGAAGAAAAACA	3449
Db	46605	CTCTAAGCTGGACGGCTTCTGGAACGGAGGATTAGACAAATTTACACTGGAAGAAAAACA	46664
QY	3450	GCACATCGAAAAAATCAAGTTGGAGGGTGAATTAAGGGTATAGGAAGAGCTTCTACAAA	3509
Db	46665	GCACATCGAAAAAATCAAGTTGGAGGGTGAATTAAGGGTATAGGAAGAGCTTCTACAAA	46724
QY	3510	TTCTTCAAAAAATCTCTGTAATCACAGTATCAACAGTAATAACGAAAAAGAGGGTGTCAAG	3569
Db	46725	TGCTGTAAAAAGTCTCTGAGCTCGGTATCCACAAAAAGCAGACGAAGGCGCACCAGG	46784
QY	3570	TGACTCGATGACAGACAG	3629
Db	46785	TGACTTGTCTTAGACAG	46844
QY	3630	TCAGGAGATGTTCAAAAGTATTCCTCAGTCTTTAGAAATCAGTGTATCCTAGTATACAC	3689
Db	46845	GCTGGGGGCTCACAGAGTATCCCTTGGTTCCTTGGCATCAGTCTCTAGTCTCAGCAC	46904
QY	3690	AAACAACTTATCCAAAAGATCGAGTGTGTAGATGTCTCCATTCGGAGCCGAGAAAC	3749
Db	46905	GGCAAACTGTATCCAAAAGTCAAGTATTAGATGTCTCCATTCAGAGTCCGGGGCC	46964
QY	3750	AAATGTCCGAAACAAATTCATTAAGATGACATGACATGACATGACATGACATGACATG	3809
Db	46965	AGCTGTACAGGGCAGAACTCTCTTGAAGCGACCTAGATGATGATGATGATGATGATG	47024
QY	3810	CAGTAGAGCCAGGAACCCACTAAGAGTAAACCAAGGAAATGATTTTTTCATCGATGA	3869
Db	47025	CAGTAGAGCCAGGAACCCACTAAGAGTAAACCAAGGAAATGATTTTTTCATCGATGA	47084
QY	3870	CTCTAACTAGCCAGTGCAGATGATATTTGGTACTTTTGTATCTGTGAAGCAAAAAACCG	3929
Db	47085	CTCTAACTAGCCAGTGCAGATGATATTTGGTACTTTTGTATCTGTGAAGCAAAAAACCG	47144
QY	3930	CATACAGGAGGAAAGTGCACACCATTTT---TTCTTCTTCCAAAGAGCTTTTACATCATC	3986
Db	47145	CCACAGGAGGATTTGAGCACTGCTGTGTCTATCTTCTTCCAAAGAGTATTTAGCTGACC	47204
QY	3987	AGTGCCTAAAAGTACCAATGACAGAGATGCCACACCTCTCTGTCAGAGCAATGAGCTTTGA	4046
Db	47205	AGTGCCTAAAAGTACCAATGACAGAGATGCCACACCTCTCTGTCAGAGCAATGAGCTTTGA	47264

Qy	4047	AGGAAACCTGGGATGTGACTCTGAATCTAATAGACACTTTGGAAATAGTTCTGTATACCGT	4106
Db	47265	TGGGAGACTAGGTGGTACTCTGAATATAATAGCACTTTGGAAATAACTCTGTCTAT	47324
Qy	4107	GTCTATTACAGATAGCAGTGAAGAGATATGATTTTCAGNATAGCAATGNAACATTC	4166
Db	47325	GTATATCGGGACACAGTGAAGAGATATGTTTTCAGNATAGCAATGACACACCTC	47384
Qy	4167	TGAACAGTTTCAGAACTCGAGAAACAGATGTTGAAGTCTTGAGGCGGTAAAGTGTGAGTT	4226
Db	47385	TAAACGGTTCATACTCCAGACCAAGCATGGAAGTAGCGGCGGCACAAATGTCAGTT	47444
Qy	4227	GGTTCTGGTGAGTCCACTGTGAACACTGTGAGGACAGCGTCCGCGTCAAGGGGACTGAAGC	4286
Db	47445	AGTTTC---TAAGTCCACTGAAATTTGTGACACAGACTACAGGGTAAAGTAACTGAAGC	47501
Qy	4287	AAATGGTAAACCAAGTCCAGCAGAGAAATTTAGAGAGAGACCACTTAATRAATGTAG	4346
Db	47502	AAATGGTAAAGAACTGGGCCAGCATCAAAAGCCAGAGGAGAGACGTTAAACAGATGTGC	47561
Qy	4347	TGATCAATAAAGCTTAAAAATACCCTGACAAAAAGAAATAATGAAAATCGAGAGTCTGA	4406
Db	47562	CGACCAAGTTAGCTTAAGACACAGCGCTGACAGAAAGAACAGTGAAGCCCGAGAGTCTGA	47621
Qy	4407	AAAGAAAGGACAGACAGCAAGTACATTTCAAAATAATGGAAAAAGATATAAACCCAAAT	4456
Db	47622	AAAGAGAGACAGAAAGCAATAAATTTCAAAATAATGGCAAGATAGTAAAGCTAAGG	47681
Qy	4467	ATATTGAAAGGTGAATCTTGAAGAAATTTCTGAGAGTAGAGTGTAGTGTGTAATGT	4526
Db	47682	GTATCTGAAGGACGGGCATGAGGATGTCTGTGATGGTAAAGTCATGAGTGGTCCGT	47741
Qy	4527	TGAACCAAGGTTAATATAATAAATAATCCCTGAGAAATGATATTAATCATTTGAC	4586
Db	47742	TGAACCAAGGTCATATATCAATAAAGTCAATCCCTGGAA--ATATCAAGTCACGTGC	47798
Qy	4587	TGTTAAGAATCTGTATAAGGCCATTCATTAATGTTGATGTCTATCATGGAAGATTTAA	4646
Db	47799	TGGGAAGGAATCTGTGTGAAGCCGTTTGTAAAGCGGACATCATTTATGGAAGATTCAG	47858
Qy	4647	TGAAGAAACAGCTCCGAAACAAATCGATTTCTGAGTTCTTCAGATGCTGAAGTAA	4706
Db	47859	TGAACAAACACTTCAGAAACCACTCGTATTCTAGTTCTTCAGATGCCAAAGGTA	47918
Qy	4707	CTACCGAGATAGCCTTGAGACCCCTGCCATCAACCAAGAGTCTGACAGTACACAGACG	4766
Db	47919	TCACCAAGATGGCTCCATACCTTACCATCCACCAAGACTCTGCCAGTACCAGGTAAT	47978
Qy	4767	CACACCTCAGATCTTGTCCAGAAAGCAATTCAGTTAATCAGGTAGAAATATGGAAT	4826
Db	47979	CATACCCCGCCCTTGTCCAGATCGGAACCTCACTTAATCAGGTAGAAGATATGGAAC	48038
Qy	4827	AGAAACCTCAGAAGTTAAGAAGTTACTTTCATCACTATTACTTCTGAAGAGGATCTAA	4886
Db	48039	TGAAGGTCAGAGGTTAAGAAGTTACCCGCTCTCCCATTTACTGTGAGAGGACTCTAA	48098
Qy	4887	TCTCAGTAATGACTTTATTGATGAAATGGTGTGCCATCAACCAAAATGAAATGTCAA	4946
Db	48099	CCTCGTAAAGAGGTCATGATGAAACGGTCTACCTTCCAGCAAGATGAGATGTCAA	48158
Qy	4947	TGGAGAAATCTAAAGAAAAACGGTCATCAGAGAGTACCAGAGTACCCTCCACAGTGGC	5006
Db	48159	TGGAGAAATCCCAAGAAAGAGGGTCAATCAGAGAGTACCACAAATGACATCCACTGTTC	48218
Qy	5007	CACAGAAATCAAACTGTGATCAGGTAGAAAAAGGGCATAAAGCAACTGTGGTTCTTC	5066
Db	48219	CACAGAAATCAAACTGTGATTAAGGTAGCAAAAGGGGACAGCAAACTGTGGTTCTTC	48278
Qy	5067	CACAGAAATTTGTCAAAATCCACTGTGCACACCACTACAACTGACCAAGCTTTC	5126
Db	48279	CACAGAAATCTGCCAGATCCACTGTGCACAACTACCACCACTGTGACAAAAGCTCTC	48338
Qy	5127	CACACCTCCACAGCGCGCAGTGGGACATCATCTCTGTAAGGAGAGAGCAAAACCGT	5186

Db	48339	CACACCTCCCGACAGACCGCATGTGGATACCGTCTCTGTAAAGGAGCAGAGTAAACCGT	48398
Qy	5187	GGTCACACACAGTGTGACAGACTCCCTGACACACCGGAGGACACTGGTTACATCTAT	5246
Db	48399	GGTCACACACAGTGTGACAGACTCCCTGAGCAGTGTGAGGAGACACTGGTGAGTCTAT	48458
Qy	5247	GACTGTGAGCAAGAGTATTTCCACACAGACAAAAGTGAATGATGAAATTTTCAAGACC	5306
Db	48459	GACAGTGTGAGCAAGTATTTCCACAGAGACCGAGTAAAGTCTATGAAGTTCTCAAGACC	48518
Qy	5307	AAAGAACTCTTCAGGTACAGCTCTGCCATCTATAGAAAAATTTCTTACCAGAGCAC	5366
Db	48519	CAAGAAACCGGTTTCAGGGACAGCCCTGCCGTCTATAGGAAGTTCTGTACCAAGACG	48578
Qy	5367	CAAGAAAGAGCATTTTGTGTTTGCCTTAATGATGACTTAAAAAAGTTGGCCGAAAAAGG	5426
Db	48579	CAAAAAGACATATTCGTGTTGCCTTAATGATGACTTGAANAAGCTGCCAGAAAAAGG	48638
Qy	5427	AATCCGAGAGTCCCTTATTTTAATACAATGCAAAACCTGCTTTGGATATATGGCCATA	5486
Db	48639	CATCCGGGAAGTCCCGTCTTAAATTAATGCAAAAGCCCGCTTGGATATCTGGCCATA	48698
Qy	5487	TCCTTCTCTAGACCGCTTTGGCATCAGTTGGAGGTATAGACTTCAGACAGTAAAGTC	5546
Db	48699	CCCTTCTCCCGACAGCAACCTTTGGCATCAGCTGGAGGTATGTACCTCAGAAATGTTCT	48758
Qy	5547	CTTAGCTGGAGTGAGCCTGTGTTACCGTTACTCTGGGCAAGTTTGAGATGGGATGAT	5606
Db	48759	CGGGCAGGGAGGTGCTTTTAAAGCGTTTGTAAAGTAAATGAGATGCCAGAGAT	48818
Qy	5607	GGCGGCCA 5614	
Db	48819	AGGGCTCA 48826	

Search completed: September 23, 2003, 15:44:36

Job time : 23092 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 09:12:14 ; Search time 1522 Seconds
(without alignments)
17496.686 Million cell updates/sec

Title: US-09-698-295-2

Perfect score: 9865
Sequence: 1 ggcacggtgagtggtgccc.....acaaaaaaaaaaaaaaaa 9865

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
- 25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9865	100.0	9865	21	Human transcriptio
2	9067	91.9	9700	21	Human transcriptio
3	7255.2	73.5	8280	25	Human nucleic acid
4	2565.6	26.0	2673	24	Human cDNA differe
5	1719.4	17.4	2057	22	Human cDNA sequenc
6	1240.2	12.6	1540	22	Human diagnostic a
7	814.6	8.3	1032	20	Human secreted pro
8	771.4	7.8	841	22	Human cDNA clone (

9	735	7.5	735	24	ABK47549	DNA encoding SCAM/
10	673.8	6.8	1408	21	AAC99129	Human pancreatic c
11	627.4	6.4	750	22	AAI94524	Human neuroblastom
12	574	5.8	574	25	ACA57390	Human adipocyte Se
13	574	5.8	2123	22	AAH17323	Human cDNA sequenc
c 14	538.2	5.5	556	22	AAH09968	Human CDNA clone (
15	413	4.2	763	20	AAZ17218	Lung small cell ca
16	359	3.6	574	24	AA561658	Human gene express
17	355.6	3.6	368	21	AAC07933	Human secreted pro
18	348.4	3.5	15037	22	AAI99283	Human excretory re
19	348.4	3.5	15037	22	AAI63633	Human kidney relat
c 20	346.2	3.5	347	21	AAC30831	Human secreted pro
c 21	316.4	3.2	341	24	ABK45859	CDNA encoding colo
c 22	303	3.1	731	22	AAI94525	Human neuroblastom
c 23	263	2.7	265	14	AAQ59932	Human brain Expres
c 24	244.8	2.5	2928	23	ABL23837	Drosophila melanog
c 25	244.8	2.5	5380	23	ABL23836	Drosophila melanog
c 26	222.8	2.3	865	23	AA583023	DNA encoding novel
c 27	220.8	2.2	247	21	AAA45848	Human secreted exp
c 28	214	2.2	214	21	AAC03141	Human secreted pro
c 29	181.4	1.8	405	24	ABS69806	Novel murine polyn
c 30	177.4	1.8	2666	24	ABK35990	CDNA sequence #281
c 31	177.4	1.8	2671	21	AAZ34559	Human secreted pro
c 32	177.4	1.8	4845	22	AA535982	Human cardiovascular
c 33	177.4	1.8	4646	22	AA535981	Human cardiovascular
c 34	176.4	1.8	411	25	ACC46298	Human dithp transc
c 35	175	1.8	452	22	AAH36921	Human colon cancer
c 36	162.8	1.7	167	16	AAT19304	Human gene signatu
c 37	157.4	1.6	311	24	ABV95834	Human pancreatic c
c 38	120.4	1.2	242	21	AAC09489	Human secreted pro
c 39	101.4	1.0	618	24	ABQ47666	Oligonucleotide fo
c 40	101.4	1.0	618	24	ABQ47667	Oligonucleotide fo
c 41	95.6	1.0	2465	11	AAQ03665	Sequence homology
c 42	91.2	0.9	589	24	ABQ50266	Oligonucleotide fo
c 43	91.2	0.9	589	24	ABQ50267	Oligonucleotide fo
c 44	90.8	0.9	157	24	ABZ08622	Human leukocyte de
c 45	90.2	0.9	360	23	ABL30495	Drosophila melanog

ALIGNMENTS

RESULT 1	
AAZ39032	
ID	AAZ39032 standard; cDNA; 9865 BP.
XX	
AC	AAZ39032;
XX	
DT	22-FEB-2000 (first entry)
XX	
DE	Human transcriptional regulatory factor TCoal gene SEQ ID NO:2.
XX	
KW	Human; transcriptional regulatory factor; TCoal; BLAST detection;
KW	bromo-domain; cell proliferation; cancer; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
CDS	Location/Qualifiers
FT	259..8982
FT	/*tag= a
FT	/product= "transcriptional regulatory factor"
XX	
PN	WO9957143-A1.
XX	
PD	11-NOV-1999.
XX	
PF	30-APR-1999; 99WO-JP02340.
XX	
PR	30-APR-1999; 98JP-0137631.
XX	
PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX	
PI	Jones MH;

XX WPI: 2000-052940/04.
DR P-PSDB: AAY57452.
XX
PT Transcriptional regulatory factor containing a bromo domain and gene
TCOA1 encoding it
XX
PS Claim 4; Page 69-98; 154pp; Japanese.
XX
CC The present sequence encodes a human transcriptional regulatory factor
CC containing a bromo domain. The factor interacts with proteins involved
CC in the chromatin-mediated transcription regulatory mechanism. It binds
CC to hSNF2H, hSNF2L and NCOA-62/Skip. It can be used for screening
CC compounds binding to it and acting as agonists or antagonists, which
CC are potentially useful for the treatment and prevention of cancer and
CC other cell proliferation disorders.
XX
SQ Sequence 9865 BP; 3338 A; 2121 C; 2218 G; 2188 T; 0 other;

Query Match 100.0%; Score 9865; DB 21; Length 9865;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9865; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGGCTGAGGTGGCGCCCAAGACGCGGCTGAGCTCGCCAGGGGGGCGCAGCTAGC 60
DB 1 GCCCAGGCTGAGGTGGCGCCCAAGACGCGGCTGAGCTCGCCAGGGGGGCGCAGCTAGC 60

QY 61 CGGAGGAAGCGGG 120
DB 61 CGGAGGAAGCGGG 120

QY 121 GGGCAGGAGGCGGGGCGGCACACGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180
DB 121 GGGCAGGAGGCGGGGCGGCACACGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180

QY 181 ACCACCGCGCGCGGAGGCGCGTCAACAAAGTGTGTACGATGACACGAGAGCGAGCGG 240
DB 181 ACCACCGCGCGCGGAGGCGCGTCAACAAAGTGTGTACGATGACACGAGAGCGAGCGG 240

QY 241 GTGGAGGAGGAGGAGGACATGGTCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
DB 241 GTGGAGGAGGAGGAGGACATGGTCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300

QY 301 GAGACCCAGGATTTCTGAGGACGACGAGGAGGATGAGATGGAAGAGGAGGAGGATGCTC 360
DB 301 GAGACCCAGGATTTCTGAGGACGACGAGGAGGATGAGATGGAAGAGGAGGAGGATGCTC 360

QY 361 GATTATCCGAGGAGATGGAAGACGACGACGACGACGACGACGACGACGACGACGACG 420
DB 361 GATTATCCGAGGAGATGGAAGACGACGACGACGACGACGACGACGACGACGACGACG 420

QY 421 TTCAGGAGCCATGATGACCTACGAGCAGCTCCAGGTAGGCGAAACCAAGAGTACATCGG 480
DB 421 TTCAGGAGCCATGATGACCTACGAGCAGCTCCAGGTAGGCGAAACCAAGAGTACATCGG 480

QY 481 COTCGTTCTCCTATATTGGAAGAAAGACATCCCGCCCCCTTGAAATTTCCCAAGTCTCT 540
DB 481 COTCGTTCTCCTATATTGGAAGAAAGACATCCCGCCCCCTTGAAATTTCCCAAGTCTCT 540

QY 541 GAGGATTTAATGGTGCCTAATGAGCATATATGATGTCATTCGCAATTTACGAGTACTG 600
DB 541 GAGGATTTAATGGTGCCTAATGAGCATATATGATGTCATTCGCAATTTACGAGTACTG 600

QY 601 CGGAATTTGGGACCTGTTTGAGATTATCTCTTTTCGCTTTGAGGACTTTTGTGCAGCT 660
DB 601 CGGAATTTGGGACCTGTTTGAGATTATCTCTTTTCGCTTTGAGGACTTTTGTGCAGCT 660

QY 661 CTGGTGAGCCAAAGACAGTGCACATCATGGCAGAGATGCATGTTGCTTTTGAAGACA 720
DB 661 CTGGTGAGCCAAAGACAGTGCACATCATGGCAGAGATGCATGTTGCTTTTGAAGACA 720

QY 721 GTTCTGCGTGAAGAGACACTTCCAAATCTACTCTTTGGACCTGCTGATCTGAAAGATAGC 780
DB 721 GTTCTGCGTGAAGAGACACTTCCAAATCTACTCTTTGGACCTGCTGATCTGAAAGATAGC 780

DB 721 GTTCTGCGTGAAGAGACACTTCCAAATCTACTCTTTGGACCTGCTGATCTGAAAGATAGC 780
QY 781 GTTAATTCACACTGTATTTTCATAGATGGGATGACGTGCCAGAGAGTCTCGGGGTGAC 840
DB 781 GTTAATTCACACTGTATTTTCATAGATGGGATGACGTGCCAGAGAGTCTCGGGGTGAC 840
QY 841 TGTGAGAGTGATAAGAGTACCATCAGTTCTTCTTACCAAGAGCGAGAGCTACCCA 900
DB 841 TGTGAGAGTGATAAGAGTACCATCAGTTCTTCTTACCAAGAGCGAGAGCTACCCA 900

QY 901 TATGGACCACTAGAGAACAGATCAAAAGTTCTACAGTTTCTAGTCGATCAGTTCTTACA 960
DB 901 TATGGACCACTAGAGAACAGATCAAAAGTTCTACAGTTTCTAGTCGATCAGTTCTTACA 960

QY 961 ACAATATTTGCTCGAGAGGAATTTGATGCTGTAAGGGGTGATACAGTATGATGATGAT 1020
DB 961 ACAATATTTGCTCGAGAGGAATTTGATGCTGTAAGGGGTGATACAGTATGATGATGAT 1020

QY 1021 AGGGTTTGTACAAACTTGGGGATTTGCTTGTGAGACATGTTTTCAGAGTATACCAT 1080
DB 1021 AGGGTTTGTACAAACTTGGGGATTTGCTTGTGAGACATGTTTTCAGAGTATACCAT 1080

QY 1081 TTGGAATGTGTGAAGCCACTCTTGAAGAGGTGCCAGAGAGAGTGGCAGTGTGAAGTC 1140
DB 1081 TTGGAATGTGTGAAGCCACTCTTGAAGAGGTGCCAGAGAGAGTGGCAGTGTGAAGTC 1140

QY 1141 TGTGTAGCACACAAGTGGCTGCTGCTGACTGACTGTTGCTGTAATCCAAAATAATAA 1200
DB 1141 TGTGTAGCACACAAGTGGCTGCTGCTGACTGACTGTTGCTGTAATCCAAAATAATAA 1200

QY 1201 CCATATATTCGACATGAACTTATGATATGATAGAGTGGAGGAAATACTTGGTCTTGG 1260
DB 1201 CCATATATTCGACATGAACTTATGATATGATAGAGTGGAGGAAATACTTGGTCTTGG 1260

QY 1261 AACCGAAGACTCATAATAGAGAAGATACAGAAATGAAATGAAAGAAATTTGGTAT 1320
DB 1261 AACCGAAGACTCATAATAGAGAAGATACAGAAATGAAATGAAAGAAATTTGGTAT 1320

QY 1321 TACAGCACAAAGTCCAACTTGCAGATTAATTTGACTGTCTAGACAAAGATTTTGGGAA 1380
DB 1321 TACAGCACAAAGTCCAACTTGCAGATTAATTTGACTGTCTAGACAAAGATTTTGGGAA 1380

QY 1381 GCAGAACTCTGCAAAATTTCTAGAGAAATCGGTGAAGAAATCCACGACACATGGACATA 1440
DB 1381 GCAGAACTCTGCAAAATTTCTAGAGAAATCGGTGAAGAAATCCACGACACATGGACATA 1440

QY 1441 ACTGAAGACCTGACCAATAAGGCTCGGGGCGAGTAAACAAATCTTTCTGGCGGCGAGTAA 1500
DB 1441 ACTGAAGACCTGACCAATAAGGCTCGGGGCGAGTAAACAAATCTTTCTGGCGGCGAGTAA 1500

QY 1501 GAAGAAATTTTGGATCCATTAAGAGCCAAAGAGGAGACATTTGATATGTTTAAAGCCCA 1560
DB 1501 GAAGAAATTTTGGATCCATTAAGAGCCAAAGAGGAGACATTTGATATGTTTAAAGCCCA 1560

QY 1561 GAAGAAACAGAAAAAGACAAAGATGAGACTGAGAAATGACTTAAAGATGCTGAGAAAAAC 1620
DB 1561 GAAGAAACAGAAAAAGACAAAGATGAGACTGAGAAATGACTTAAAGATGCTGAGAAAAAC 1620

QY 1621 AGAAGAAATTTGAAGCCAGTCCCTTGAAGACAGTGCAGCACAACCAACAGATGAT 1680
DB 1621 AGAAGAAATTTGAAGCCAGTCCCTTGAAGACAGTGCAGCACAACCAACAGATGAT 1680

QY 1681 GACCTGAGCAAGGAAATCTGAGAGCCAAACAGAGTTGGGATTAAGGTAAGTCTCTGTG 1740
DB 1681 GACCTGAGCAAGGAAATCTGAGAGCCAAACAGAGTTGGGATTAAGGTAAGTCTCTGTG 1740

QY 1741 TCAGCAAACTTTGGCGCAACACAAACAAATGCAACTTTCAGAAAGAGACTAGTCCCTCTGAA 1800
DB 1741 TCAGCAAACTTTGGCGCAACACAAACAAATGCAACTTTCAGAAAGAGACTAGTCCCTCTGAA 1800

QY 1801 GGGAGGAGCCCTGTGGGGTGTCTCTCAGAAACCCCGCATAGCAGCAACATGGCAGAGAAG 1860
DB 1801 GGGAGGAGCCCTGTGGGGTGTCTCTCAGAAACCCCGCATAGCAGCAACATGGCAGAGAAG 1860

QY	1861	AAGGTGGCATCTGAGTCCCTCCCGAGGATGTCAGAGAACCTTAACAAGACATGTGAGAGC	1920	2941	TGGAGCTGGATTAGTAAAACTCATGTTTATAGTGTTCCTAAATGGCAGGCAATACT	3000
Db	1861	AAGGTGGCATCTGAGTCCCTCCCGAGGATGTCAGAGAACCTTAACAAGACATGTGAGAGC	1920	2941	TGGAGCTGGATTAGTAAAACTCATGTTTATAGTGTTCCTAAATGGCAGGCAATACT	3000
QY	1921	AGTAACACTAGTGTCTACCACTACCTCCATCCAGCCTAATCTCGAAGAACAGTAAACAGCAGC	1980	3001	AATGTGAATTACAGAAAGTCGTTAGAGGAACCAAAAATAATATGATGAAATATGGAT	3060
Db	1921	AGTAACACTAGTGTCTACCACTACCTCCATCCAGCCTAATCTCGAAGAACAGTAAACAGCAGC	1980	3001	AATGTGAATTACAGAAAGTCGTTAGAGGAACCAAAAATAATATGATGAAATATGGAT	3060
QY	1981	AGTGAACATAATCTTCCAGAGTGAATCTGTAAGGAGCTGATGATCCTGAAATGGA	2040	3061	GAGTCAGATATAAGAAAAATGTTTACGAAAGTCACAAAAATAAATAAGAGCCTGATCT	3120
Db	1981	AGTGAACATAATCTTCCAGAGTGAATCTGTAAGGAGCTGATGATCCTGAAATGGA	2040	3061	GAGTCAGATATAAGAAAAATGTTTACGAAAGTCACAAAAATAAATAAGAGCCTGATCT	3120
QY	2041	GAAAGAGATCTCATACACTGCTCTATCTAGGAGAGATAGTAGGTAATTCACATCG	2100	3121	GAAAAAGATGAGTAAAGAGTTTACGAAAGTCGAAAGAGGACACCAAAATGAAATGGAT	3180
Db	2041	GAAAGAGATCTCATACACTGCTCTATCTAGGAGAGATAGTAGGTAATTCACATCG	2100	3121	GAAAAAGATGAGTAAAGAGTTTACGAAAGTCGAAAGAGGACACCAAAATGAAATGGAT	3180
QY	2101	GAGAGTCCACCGGGAGCTAAGTGAATCTCTGAGCTGGAAGAGGACATCTGGCTCA	2160	3181	ATCTCAAGATTAATCTAGAGAGGACCAAGATGTGAAGAGCTCTTAGATTCTGACAGT	3240
Db	2101	GAGAGTCCACCGGGAGCTAAGTGAATCTCTGAGCTGGAAGAGGACATCTGGCTCA	2160	3181	ATCTCAAGATTAATCTAGAGAGGACCAAGATGTGAAGAGCTCTTAGATTCTGACAGT	3240
QY	2161	ACTCGAATCATCACCAGATTCGGGAATCCAGATAGCAAACTTATGTCAGTGAAGAGCCAG	2220	3241	GATAAACCTCTGCAAGGAAGAACCAATGGAAGTAGACGATGACATGAAAAACAGATCACAT	3300
Db	2161	ACTCGAATCATCACCAGATTCGGGAATCCAGATAGCAAACTTATGTCAGTGAAGAGCCAG	2220	3241	GATAAACCTCTGCAAGGAAGAACCAATGGAAGTAGACGATGACATGAAAAACAGATCACAT	3300
QY	2221	CAGGTGGCAGCGCTGCACATGAAGCAATAAATATTTAAGGAGGGCAAGAGGTACTG	2280	3301	GTAAATTTGTCCAGAGAGTTTCTCAAGTAGATGTGGTCAATGTAGTCAGGGTTTTCATCTA	3360
Db	2221	CAGGTGGCAGCGCTGCACATGAAGCAATAAATATTTAAGGAGGGCAAGAGGTACTG	2280	3301	GTAAATTTGTCCAGAGAGTTTCTCAAGTAGATGTGGTCAATGTAGTCAGGGTTTTCATCTA	3360
QY	2281	GTAGTTAACTCTCAAGGAGAAATTTACGGTTGAGCACCACCAAAAGAGTATCATGAA	2340	3361	AGGACTAGTTTACAAAAGAAAAACAATCATCCAACTAGATGGAGCTCTTGAAGAGGA	3420
Db	2281	GTAGTTAACTCTCAAGGAGAAATTTACGGTTGAGCACCACCAAAAGAGTATCATGAA	2340	3361	AGGACTAGTTTACAAAAGAAAAACAATCATCCAACTAGATGGAGCTCTTGAAGAGGA	3420
QY	2341	GGAAATATCAACAATATTTAAATTTGGGTCAAGAGGAAGTATCGCGTCTACACAAAT	2400	3421	ATTAAACAGTTTACACTGGAAGAAAAACAGCGACTCGAAAAATCAAGTTGGAGGTGGA	3480
Db	2341	GGAAATATCAACAATATTTAAATTTGGGTCAAGAGGAAGTATCGCGTCTACACAAAT	2400	3421	ATTAAACAGTTTACACTGGAAGAAAAACAGCGACTCGAAAAATCAAGTTGGAGGTGGA	3480
QY	2401	CAATACTCCACAAATTCATTTGCTTTGAATAGCACCAGCACAGAGAACCATGATAAG	2460	3481	ATTAGGGTATAGGAAAGACTTCTCAAAATTCCTTCAAAAATCTCTCTGAATCACCAGTA	3540
Db	2401	CAATACTCCACAAATTCATTTGCTTTGAATAGCACCAGCACAGAGAACCATGATAAG	2460	3481	ATTAGGGTATAGGAAAGACTTCTCAAAATTCCTTCAAAAATCTCTCTGAATCACCAGTA	3540
QY	2461	AGAAGGCATCTTGCATAGTGTAGCAAACTCTGCTGACTCCAGCAGGAGGTTCAATGGACGGT	2520	3541	ATAAGCAAGCAAAAGAGGGTGTCTAGAGTGTCTGAGAGTGTCTGAGCAACAAAGTCCCAAT	3600
Db	2461	AGAAGGCATCTTGCATAGTGTAGCAAACTCTGCTGACTCCAGCAGGAGGTTCAATGGACGGT	2520	3541	ATAAGCAAGCAAAAGAGGGTGTCTAGAGTGTCTGAGAGTGTCTGAGCAACAAAGTCCCAAT	3600
QY	2521	TCTGTCCATGGTCCAAAGTCTTACCATATCTACTCTGAGACTGACTATACCCAAATTA	2580	3601	GCAATAATATGATCAACCTGAGGACTTTGATTGAGGGATGTTTCAAAAAGTATTCTCCTCAGTT	3660
Db	2521	TCTGTCCATGGTCCAAAGTCTTACCATATCTACTCTGAGACTGACTATACCCAAATTA	2580	3601	GCAATAATATGATCAACCTGAGGACTTTGATTGAGGGATGTTTCAAAAAGTATTCTCCTCAGTT	3660
QY	2581	GAAACACATCCCTTCATCTTTTTCATCCCACTGGGCAATCATAGGCAATTTGG	2640	3661	CTTAGAATGAGTGTCTAGTCTATACCAACAAACTTTATCCAAAAGATCGAGTGTTA	3720
Db	2581	GAAACACATCCCTTCATCTTTTTCATCCCACTGGGCAATCATAGGCAATTTGG	2640	3661	CTTAGAATGAGTGTCTAGTCTATACCAACAAACTTTATCCAAAAGATCGAGTGTTA	3720
QY	2641	ATCAAGGCAGTTTCAGATGTAGCAAACTCCAGAGTGTGCAATGGCTTTAGCCATTTTG	2700	3721	GATGATGTCTCCATTCGGAGCCAGAAAAACAATGTCCGAAAAACAATTTCCATTGAAAAT	3780
Db	2641	ATCAAGGCAGTTTCAGATGTAGCAAACTCCAGAGTGTGCAATGGCTTTAGCCATTTTG	2700	3721	GATGATGTCTCCATTCGGAGCCAGAAAAACAATGTCCGAAAAACAATTTCCATTGAAAAT	3780
QY	2701	GAGTGTGAGTTTAAACCACTGCTGATGCTTACCAATATGGCGAGATTTTGGACATACC	2760	3781	GACATAGAGAAAAAGTCTCTGACCTTGCAGTAGAGGCCAGGAACCCACTAAGAGTAAA	3840
Db	2701	GAGTGTGAGTTTAAACCACTGCTGATGCTTACCAATATGGCGAGATTTTGGACATACC	2760	3781	GACATAGAGAAAAAGTCTCTGACCTTGCAGTAGAGGCCAGGAACCCACTAAGAGTAAA	3840
QY	2761	AGGTTTACCCGGATGACATCAATGAAAGAGAGAAAAAGGAGAAAGTCAAAAAAAGAG	2820	3841	ACCAAGGAAATGATTTTTTTCATCGATGACTCTAAACTAGCCAGTGCAGATGATTTGGT	3900
Db	2761	AGGTTTACCCGGATGACATCAATGAAAGAGAGAAAAAGGAGAAAGTCAAAAAAAGAG	2820	3841	ACCAAGGAAATGATTTTTTTCATCGATGACTCTAAACTAGCCAGTGCAGATGATTTGGT	3900
QY	2821	AGAACAGGAGAGAGAAACGATGACGAGGACATGGGTAAAATACACATTTCCA	2880	3901	ACTTTTGTCTGTAAGAACAAAAACCGCTCATACAGGAGGAAAGTGACACCATTTGTTCT	3960
Db	2821	AGAACAGGAGAGAGAAACGATGACGAGGACATGGGTAAAATACACATTTCCA	2880	3901	ACTTTTGTCTGTAAGAACAAAAACCGCTCATACAGGAGGAAAGTGACACCATTTGTTCT	3960
QY	2881	GTAAAGCATCAGTTTGGAAACAAAAAGGTGAAGAGTACAGAGTACAGGATATGTTGTT	2940	3961	TCCTTCCAGAGTGTCTTACATTCATCAGTGCCTTAAAGTACCAATGACAGAGATGCCACA	4020
Db	2881	GTAAAGCATCAGTTTGGAAACAAAAAGGTGAAGAGTACAGAGTACAGGATATGTTGTT	2940	3961	TCCTTCCAGAGTGTCTTACATTCATCAGTGCCTTAAAGTACCAATGACAGAGATGCCACA	4020
				4021	CCTCTGTCAAGAGCAATGGACTTTGAAGGAAAACTGGGATGTGACTCTCTGAATCTAATAGC	4080

QY 6241 GTCCTGGGTATCATTTCCATCAAGTACAGGTACCGTACAGTACAGCAAACTTTTACTTTCATTCOAG 6300
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6241 GTCCTGGGTATCATTTCCATCAAGTACAGGTACCGTACAGTACAGCAAACTTTTACTTTCATTCOAG 6300
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6301 CCCAGGACAGCAAGTACAGTACAAATTAGGCCCAATACCTTACAGTCTGGAGGAAACACAGC 6360
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6301 CCCAGGACAGCAAGTACAGTACAAATTAGGCCCAATACCTTACAGTCTGGAGGAAACACAGC 6360
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6361 AATTCAAGTAATACAGAGGCTCAGATTGCCCTTGTTATGACCGTATTAGAACACCA 6420
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6361 AATTCAAGTAATACAGAGGCTCAGATTGCCCTTGTTATGACCGTATTAGAACACCA 6420
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6421 CTCGAACAGTCAACACTAGGAAGGCAATATTGCAACACCTGTGATGGTACAGCCAGGT 6480
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6421 CTCGAACAGTCAACACTAGGAAGGCAATATTGCAACACCTGTGATGGTACAGCCAGGT 6480
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6481 GCTCTCAGCAAGTACACTCAATATCATCAGGGGGAGCGCTGCTCCACTGCAGTCTCC 6540
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6481 GCTCTCAGCAAGTACACTCAATATCATCAGGGGGAGCGCTGCTCCACTGCAGTCTCC 6540
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6541 GCCCTTAACACGGTTTCTCAACACCTGGGCGAGAAAGCTTTAACTTCAGCAACGTCCTACT 6600
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6541 GCCCTTAACACGGTTTCTCAACACCTGGGCGAGAAAGCTTTAACTTCAGCAACGTCCTACT 6600
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6601 TCAATATACAGTCTTACAGCTCACAACCCCTCGCCCCCAACAAAGGACAAAGTGAAGCTC 6660
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6601 TCAATATACAGTCTTACAGCTCACAACCCCTCGCCCCCAACAAAGGACAAAGTGAAGCTC 6660
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6661 ACCATGGCTCAACTTACTCAGTTTACACAGGGCCACGGTGCAATCAAGTTTGCAGAGTA 6720
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6661 ACCATGGCTCAACTTACTCAGTTTACACAGGGCCACGGTGCAATCAAGTTTGCAGAGTA 6720
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6721 GTAATTCAAGGACAAAGTCAAACTACTGGACAGTTGCAAGTTGATACCTCAAGGGGTGACT 6780
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6721 GTAATTCAAGGACAAAGTCAAACTACTGGACAGTTGCAAGTTGATACCTCAAGGGGTGACT 6780
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6781 GTACTCCAGGCCAGGCCAGCGTAATGCAAGTGCAATGCGCAATGCGTACTGTTCCAG 6840
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6781 GTACTCCAGGCCAGGCCAGCGTAATGCAAGTGCAATGCGCAATGCGTACTGTTCCAG 6840
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6841 CGATTCTCTTTTACCCATTGGCAACACAGCCACACAGCCAGCCAGCCAGCCAGCCACT 6900
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6841 CGATTCTCTTTTACCCATTGGCAACACAGCCACACAGCCAGCCAGCCAGCCAGCCACT 6900
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6901 GTTTCCAGCAGCAGCAGGTACAGGTGAACAAAGGCGAGAGTAACTGTCAACCCAGATG 6960
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6901 GTTTCCAGCAGCAGCAGGTACAGGTGAACAAAGGCGAGAGTAACTGTCAACCCAGATG 6960
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6961 CAGGTACATCAAGCAAAACCTGCGCAGCTCAGTCAATCAAGTGTGGGTCCAGCAAAA 7020
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6961 CAGGTACATCAAGCAAAACCTGCGCAGCTCAGTCAATCAAGTGTGGGTCCAGCAAAA 7020
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7021 GCCCAGCCACAGTCTGCTCAGCTTACGCTGGCCCCAGCCCCCAACCCAGCCAGTCC 7080
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7021 GCCCAGCCACAGTCTGCTCAGCTTACGCTGGCCCCAGCCCCCAACCCAGCCAGTCC 7080
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7081 CCAGCTCAGCTGAAGTTTCAAGTTCAGAGTTCAGAGCCCAAACTGTTTCACTCC 7140
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7081 CCAGCTCAGCTGAAGTTTCAAGTTCAGAGTTCAGAGCCCAAACTGTTTCACTCC 7140
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7141 CATGTCCCTTCTGAAGCAACACCCAGCCAGTCAATCCAGCCCAAGTGTGACGA 7200
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7141 CATGTCCCTTCTGAAGCAACACCCAGCCAGTCAATCCAGCCCAAGTGTGACGA 7200
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7201 CAGTCTCAGCTCAAGTAAATGTTCCAGGAGCAGTCTCCTGTGTCCTCAAGTCCATCA 7260
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7201 CAGTCTCAGCTCAAGTAAATGTTCCAGGAGCAGTCTCCTGTGTCCTCAAGTCCATCA 7260
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7261 CAGACTCGAATACGTTCCATCAACTCCATCCCAACTGTCTCCTGGCAACAAATCCAGGTT 7320
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7261 CAGACTCGAATACGTTCCATCAACTCCATCCCAACTGTCTCCTGGCAACAAATCCAGGTT 7320

QY 7321 CAGACTACAACCTCACAACCGATTCCAAATTCACAACACATACATCTCTCTCAGATACCTTCC 7380
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7321 CAGACTACAACCTCACAACCGATTCCAAATTCACAACACATACATCTCTCTCAGATACCTTCC 7380
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7381 CAAGGCCAGCCACAGTCAACAACCCAGGTACAGTCTTCAACTCAAACTCTTTCATCAGGA 7440
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7381 CAAGGCCAGCCACAGTCAACAACCCAGGTACAGTCTTCAACTCAAACTCTTTCATCAGGA 7440
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7441 CAAACTTTAAATCAAGTTAGTGTTCATCCCTCCCATCCCGTCCCTCAGCTCAAAATACAGCAG 7500
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7441 CAAACTTTAAATCAAGTTAGTGTTCATCCCTCCCATCCCGTCCCTCAGCTCAAAATACAGCAG 7500
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7501 CCAGAGCCCCAAGTCATTGTGTGCTCAGCTGCAACAACAAAGTCCAGTCTCTCTCAG 7560
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7501 CCAGAGCCCCAAGTCATTGTGTGCTCAGCTGCAACAACAAAGTCCAGTCTCTCTCAG 7560
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7561 ATCCAGTCCACAGGTTGTGCTCAGATACAGGCTCAGAAAGTGTGTCGCCAGCAAAATC 7620
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7561 ATCCAGTCCACAGGTTGTGCTCAGATACAGGCTCAGAAAGTGTGTCGCCAGCAAAATC 7620
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7621 AAACCTCCAGTTACCTTATCCAAATTCAGCAAAAGCAGTGTGTGCAGACTCACAGATTTCAG 7680
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7621 AAACCTCCAGTTACCTTATCCAAATTCAGCAAAAGCAGTGTGTGCAGACTCACAGATTTCAG 7680
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7681 AATGTGGTTACAGTGCAGGCGCCAGTGTGCAAGAGCAGTTGCAAAAGGTTTCAGCAACTC 7740
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7681 AATGTGGTTACAGTGCAGGCGCCAGTGTGCAAGAGCAGTTGCAAAAGGTTTCAGCAACTC 7740
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7741 AGGGATCAGCAGCAAAAAGAAACAGCAACAGATAGAAATTAAGCTGAACACACCCCTCC 7800
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7741 AGGGATCAGCAGCAAAAAGAAACAGCAACAGATAGAAATTAAGCTGAACACACCCCTCC 7800
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7801 AAGCTTCTTAATCAAAAGTTGAAATCATTCAGAAACAGTGGTGTGATGAAGCATTAATGCTGTA 7860
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7801 AAGCTTCTTAATCAAAAGTTGAAATCATTCAGAAACAGTGGTGTGATGAAGCATTAATGCTGTA 7860
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7861 ATAGAACATTTAAACAGAAAGAGCATGACTCCAGCTGAAAGAGAGAGAAATCAAGA 7920
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7861 ATAGAACATTTAAACAGAAAGAGCATGACTCCAGCTGAAAGAGAGAGAAATCAAGA 7920
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7921 ATGATGTCTGTAAACAGTGTATGAAGTATATTTTGGATAGATAGATAAGAAAGAA 7980
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7921 ATGATGTCTGTAAACAGTGTATGAAGTATATTTTGGATAGATAGATAAGAAAGAA 7980
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7981 CAGCAGCAAAAACGGAAGCGTGAAGAGAGTGTGGAGCAGAAAACGTAGCAAGCAGAT 8040
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
7981 CAGCAGCAAAAACGGAAGCGTGAAGAGAGTGTGGAGCAGAAAACGTAGCAAGCAGAT 8040
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8041 GCCACTAAGCTGTAGCTCTGCTCTTTCAAGCAAAAGAGCAGCTCAGAGCCGAGATCCTG 8100
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8041 GCCACTAAGCTGTAGCTCTGCTCTTTCAAGCAAAAGAGCAGCTCAGAGCCGAGATCCTG 8100
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8101 AAGAGAGAGCAGCTCTCGCAAGGATCTGCAAAATGAAGTGCAGAGAGAGTGAAGAGA 8160
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8101 AAGAGAGAGCAGCTCTCGCAAGGATCTGCAAAATGAAGTGCAGAGAGAGTGAAGAGA 8160
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8161 GACCTGAAAATTAAGAAAGAAAAGACCTGTATGCAAGTTGGCTCAGGCCACAGCAGTAGCT 8220
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8161 GACCTGAAAATTAAGAAAGAAAAGACCTGTATGCAAGTTGGCTCAGGCCACAGCAGTAGCT 8220
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8221 GCACCTGCCCTCCAGTGACACAGTCTTTCAGCCCTCCAGCCCTTCCAGCTTCCAGCT 8280
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8221 GCACCTGCCCTCCAGTGACACAGTCTTTCAGCCCTCCAGCTTCCAGCTTCCAGCT 8280
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8281 CCCCTCCACCTGGTGTCAACACAGSCCTTCTGTCACGCCCTTACCTTACCTTGTCT 8340
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8281 CCCCTCCACCTGGTGTCAACACAGSCCTTCTGTCACGCCCTTACCTTACCTTGTCT 8340
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8341 TCCCAGAGAGAGGAGCGGGAAGAGAAAAGACTCCAGCTCAAAAGTCCAAAGAAAAAGAAA 8400
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8341 TCCCAGAGAGAGGAGCGGGAAGAGAAAAGACTCCAGCTCAAAAGTCCAAAGAAAAAGAAA 8400
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
8401 ATGATCTCTACTACTCAAGGAAAGAACTAAGAAAGGACAAAAGCTTTTACTGTATCTGTA 8460

RESULT 2

AAZ39033;

human transcriptional regulatory factor TCoA1 gene

Homo sapiens.

CDS
4/2..8817

W09957143-A1

◀
◀
◀
▶

|
|
|

◻
◻

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Jones MH;

1000

67 : A 100 130. 154 pp. Tananoo

The present sequence encodes a human transcriptional regulatory factor containing a bromo domain. The factor interacts with proteins involved in the chromatin-mediated transcription regulatory mechanism. It binds to hSNF2H, hSNF2L and NCOA-62/Skip. It can be used for screening compounds binding to it and acting as agonists or antagonists, which are potentially useful for the treatment and prevention of cancer and

CC other cell proliferation disorders.

Sequence 9700 BP; 3235 A; 2127 C; 2199 G; 2139 T; 0 other;

Query Match 91.9%; Score 9067; DB 21; Length 9700;
Best Local Similarity 96.1%; Pred. No. 0;
Matches 9480; Conservative 0; Mismatches 5; Indels 380;

[illegible]

1918 -----CTAGGTGATTTCAAATCGG 1936
Db
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2659
2660
2661
2662
2663
2664
2665
2666
2667
2668
2669
2670
2671
2672
2673
2674
2675
2676
2677
2678
2679
2680
2681
2682
2683
2684
2685
2686
2687
2688
2689
2690
2691
2692
2693
2694
2695
2696
2697
2698
2699
2700
2701
2702
2703
2704
2705
2706
2707
2708
2709
2710
2711
2712
2713
2714
2715
2716
2717
2718
2719
2720
2721
2722
2723
2724
2725
2726
2727
2728
2729
2730
2731
2732
2733
2734
2735
2736
2737
2738
2739
2740
2741
2742
2743
2744
2745
2746
2747
2748
2749
2750
2751
2752
2753
2754
2755
2756
2757
2758
2759
2760
2761
2762
2763
2764
2765
2766
2767
2768
2769
2770
2771
2772
2773
2774
2775
2776
2777
2778
2779
2780
2781
2782
2783
2784
2785
2786
2787
2788
2789
2790
2791
2792
2793
2794
2795
2796
2797
2798
2799
2800
2801
2802
2803
2804
2805
2806
2807
2808
2809
2810
2811
2812
2813
2814
2815
2816
2817
2818
2819
2820
2821
2822
2823
2824
2825
2826
2827
2828
2829
2830
2831
2832
2833
2834
2835
2836
2837
2838
2839
2840
2841
2842
2843
2844
2845
2846
2847
2848
2849
2850
2851
2852
2853
2854
2855
2856
2857
2858
2859
2860
2861
2862
2863
2864
2865
2866
2867
2868
2869
2870
2871
2872
2873
2874
2875
2876
2877
2878
2879
2880
2881
2882
2883
2884
2885
2886
2887
2888
2889
2890
2891
2892
2893
2894
2895
2896
2897
2898
2899
2900
2901
2902
2903
2904
2905
2906
2907
2908
2909
2910
2911
2912
2913
2914
2915
2916
2917
2918
2919
2920
2921
2922
2923
2924
2925
2926
2927
2928
2929
2930
2931
2932
2933
2934
2935
2936
2937
2938
2939
2940
2941
2942
2943
2944
2945
2946
2947
2948
2949
2950
2951
2952
2953
2954
2955
2956
2957
2958
2959
2960
2961
2962
2963
2964
2965
2966
2967
2968
2969
2970
2971
2972
2973
2974
2975
2976
2977
2978
2979
2980
2981
2982
2983
2984
2985
2986
2987
2988
2989
2990
2991
2992
2993
2994
2995
2996
2997
2998
2999
3000
3001
3002
3003
3004
3005
3006
3007
3008
3009
3010
3011
3012
3013
3014
3015
3016
3017
3018
3019
3020
3021
3022
3023
3024
3025
3026
3027
3028
3029
3030
3031
3032
3033
3034
3035
3036
3037
3038
3039
3040
3041
3042
3043
3044
3045
3046
3047
3048
3049
3050
3051
3052
3053
3054
3055
3056
3057
3058
3059
3060
3061
3062
3063
3064
3065
3066
3067
3068
3069
3070
3071
3072
3073
3074
3075
3076
3077
3078
3079
3080
3081
3082
3083
3084
3085
3086
3087
3088
3089
3090
3091
3092
3093
3094
3095
3096
3097
3098
3099
3100
3101
3102
3103
3104
3105
3106
3107
3108
3109
3110
3111
3112
3113
3114
3115
3116
3117
3118
3119
3120
3121
3122
3123
3124
3125
3126
3127
3128
3129
3130
3131
3132
3133
3134
3135
3136
3137
3138
3139
3140
3141
3142
3143
3144
3145
3146
3147
3148
3149
3150
3151
3152
3153
3154
3155
3156
3157
3158
3159
3160
3161
3162
3163
3164
3165
3166
3167
3168
3169
3170
3171
3172
3173
3174
3175
3176
3177
3178
3179
3180
3181
3182
3183
3184
3185
3186
3187
3188
3189
3190
3191
3192
3193
3194
3195
3196
3197
3198
3199
3200
3201
3202
3203
3204
3205
3206
3207
3208
3209
3210
3211
3212
3213
3214
3215
3216
3217
3218
3219
3220
3221
3222
3223
3224
3225
3226
3227
3228
3229
3230
3231
3232
3233
3234
3235
3236
3237
3238
3239
3240
3241
3242
3243
3244
3245
3246
3247
3248
3249
3250
3251
3252
3253
3254
3255
3256
3257
3258
3259
3260
3261
3262
3263
3264
3265
3266
3267
3268
3269
3270
3271
3272
3273
3274
3275
3276
3277
3278
3279
3280
3281
3282
3283
3284
3285
3286
3287
3288
3289
3290
3291
3292
3293
3294
3295
3296
3297
3298
3299
3300
3301
3302
3303
3304
3305
3306
3307
3308
3309
3310
3311
3312
3313
3314
3315
3316
3317
3318
3319
3320
3321
3322
3323
3324
3325
3326
3327
3328
3329
3330
3331
3332
3333
3334
3335
3336
3337
3338
3339
3340
3341
3342
3343
3344
3345
3346
3347
3348
3349
3350
3351
3352
3353
3354
3355
3356
3357
3358
3359
3360
3361
3362
3363
3364
3365
3366
3367
3368
3369
3370
3371
3372
3373
3374
3375
3376
3377
3378
3379
3380
3381
3382
3383
3384
3385
3386
3387
3388
3389
3390
3391
3392
3393
3394
3395
3396
3397
3398
3399
3400
3401
3402
3403
3404
3405
3406
3407
3408
3409
3410
3411
3412
3413
3414
3415
3416
3417
3418
3419
3420
3421
3422
3423
3424
3425
3426
3427
3428
3429
3430
3431
3432
3433
3434
3435
3436
3437
3438
3439
3440
3441
3442
3443
3444
3445
3446
3447
3448
3449
3450
3451
3452
3453
3454
3455
3456
3457
3458
3459
3460
3461
3462
3463
3464
3465
3466
3467
3468
3469
3470
3471
3472
3473
3474
3475
3476
3477
3478
3479
3480
3481
3482
3483
3484
3485
3486
3487
3488
3489
3490
3491
3492
3493
3494
3495
3496
3497
3498
3499
3500
3501
3502
3503
3504
3505
3506
3507
3508
3509
3510
3511
3512
3513
3514
3515
3516
3517
3518
3519
3520
3521
3522
3523
3524
3525
3526
3527
3528
3529
3530
3531
3532
3533
3534
3535
3536
3537
3538
3539
3540
3541
3542
3543
3544
3545
3546
3547
3548
3549
3550
3551
3552
3553
3554
3555
3556
3557
3558
3559
3560
3561
3562
3563
3564
3565
3566
3567
3568
3569
3570
3571
3572
3573
3574
3575
3576
3577
3578
3579
3580
3581
3582
3583
3584
3585
3586
3587
3588
3589
3590
3591
3592
3593
3594
3595
3596
3597
3598
3599
3600
3601
3602
3603
3604
3605
3606
3607
3608
3609
3610
3611
3612
3613
3614
3615
3616
3617
3618
3619
3620
3621
3622
3623
3624
3625
3626
3627
3628
3629
3630
3631
3632
3633
3634
3635
3636
3637
3638
3639
3640
3641
3642
3643
3644
3645
3646
3647
3648
3649
3650
3651
3652
3653
3654
3655
3656
3657
3658
3659
3660
3661
3662
3663
3664
3665
3666
3667
3668
3669
3670
3671
3672
3673
3674
3675
3676
3677
3678
3679
3680
3681
3682
3683
3684
3685
3686
3687
3688
3689
3690
3691
3692
3693
3694
3695
3696
3697
3698
3699
3700
3701
3702
3703
3704
3705
3706
3707
3708
3709
3710
3711
3712
3713
3714
3715
3716
3717
3718
3719
3720
3721
3722
3723
3724
3725
3726
3727
3728
3729
3730
3731
3732
3733
3734
3735
3736
3737
3738
3739
3740
3741
3742
3743
3744
3745
3746
3747
3748
3749
3750
3751
3752
3753
3754
3755
3756
3757
3758
3759
3760
3761
3762
3763
3764
3765
3766
3767
3768
3769
3770
3771
3772
3773
3774
3775
3776
3777
3778
3779
3780
3781
3782
3783
3784
3785
3786
3787
3788
3789
3790
3791
3792
3793
3794
3795
3796
3797
3798
3799
3800
3801
3802
3803
3804
3805
3806
3807
3808
3809
3810
3811
3812
3813
3814
3815
3816
3817
3818
3819
3820
3821
3822
3823
3824
3825
3826
3827
3828
3829
3830
3831
3832
3833
3834
3835
3836
3837
3838
3839
3840
3841
3842
3843
3844
3845
3846
3847
3848
3849
3850
3851
3852
3853
3854
3855
3856
3857
3858
3859
3860
3861
3862
3863
3864
3865
3866
3867
3868
3869
3870
3871
3872
3873
3874
3875
3876
3877
3878
3879
3880
3881
3882
3883
3884
3885
3886
3887
3888
3889
3890
3891
3892
3893
3894
3895
3896
3897
3898
3899
3900
3901
3902
3903
3904
3905
3906
3907
3908
3909
3910
3911
3912
3913
3914
3915
3916
3917
3918
3919
3920
3921
3922
3923
3924
3925
3926
3927
3928
3929
3930
3931
3932
3933
3934
3935
3936
3937
3938
3939
3940
3941
3942
3943
3944
3945
3946
3947
3948
3949
3950
3951
3952
3953

Qy	4262	GGCTGCGGTCAGGGGACCTGAAGCAAAATGGTAAACCAACAGTCAAGCAAGAAATTTAG	4321		Db	5177	ATAGAAAATTTGTTACCAAGACAGCACCAAGAAGAGCATTTTGTGTTTGGCCTAAATGATGACT	5233	
Db	4097	GGCTGCGGTCAGGGGACCTGAAGCAAAATGGTAAACCAACAGTCAAGCAAGAAATTTAG	4156		Qy	5402	TAAAAAAGTTGGCCCGAAAGAGGAGGAATCCGACAGGTCCCTTATTTTAAATTACAATGCAA	5461	
Qy	4322	AGGAGACACAGTTAATAAATGTAGTCATCAATTAAGCTTAAAAATACCACCTGCACAAA	4381		Db	5237	TAAAAAAGTTGGCCCGAAAGAGGAGGAATCCGACAGGTCCCTTATTTTAAATTACAATGCAA	5296	
Db	4157	AGGAGACACAGTTAATAAATGTAGTCATCAATTAAGCTTAAAAATACCACCTGCACAAA	4216		Qy	5462	AACCTGCTTTGGGATATATGGCCATATCTTCTCTAGACCGACCTTTGGGCATCAGTTGGA	5521	
Qy	4382	AGATAATGAAAATTCGAGAGTCTGAAAGAAAGGACAGAGAAAGTACATTTCAAAATA	4441		Db	5297	AACCTGCTTTGGGATATATGGCCATATCTTCTCTAGACCGACCTTTGGGCATCAGTTGGA	5356	
Db	4217	AGAATAATGAAAATTCGAGAGTCTGAAAGAAAGGACAGAGAAAGTACATTTCAAAATA	4276		Qy	5522	GGTATAGACTTCAGACAGATAAGTCTTACTGGAGTGAGCCTGTATGTTACGGTTFACTGT	5581	
Qy	4442	ATGAAAAGATATAAACCACCAATATATTGTAAGAGTGAATGCTGTGAAAGAAATTTCTG	4501		Db	5357	GGTATAGACTTCAGACAGATAAGTCTTACTGGAGTGAGCCTGTATGTTACGGTTFACTGT	5416	
Db	4277	ATGAAAAGATATAAACCACCAATATATTGTAAGAGTGAATGCTGTGAAAGAAATTTCTG	4336		Qy	5582	GGCAAGTTTGGATGGGATGATATGGCGCCCAAGGTTCCTCCAGAGAGGAGGAGTACAC	5641	
Qy	4502	AGACTAGTAGTAAGTGGTAAATGTTGAAACCAAGGTTAATAATATAAATAAATATATCC	4561		Db	5417	GGCAAGTTTGGATGGGATGATATGGCGCCCAAGGTTCCTCCAGGAGGAGGAGTACAC	5476	
Db	4337	AGAGTAGTAGTAAGTGGTAAATGTTGAAACCAAGGTTAATAATATAAATAAATATATCC	4396		Qy	5642	GGACAGAAACATCCGAAACTGAAATCACACACACAGAAATAATTAAGAGAGAGAGATGTTG	5701	
Qy	4562	CTGAGAATGATATTAATCATCTGACTGTTAAAGAAATCTGTATAAGGCCATTCATTAAATG	4621		Db	5477	GGACAGAAACATCCGAAACTGAAATCACACACACAGAAATAATTAAGAGAGAGAGATGTTG	5536	
Db	4397	CTGAGAATGATATTAATCATCTGACTGTTAAAGAAATCTGTATAAGGCCATTCATTAAATG	4456		Qy	5702	GTCTTTATGGCATTCGATTTGATTTGATATGATACAGGAAATCATTTGTCCCATTGGAGTTTC	5761	
Qy	4622	GTGATGTCATCGGAAGATTTTAATGAAGAAACAGCTCCGAAACAAAATTCGATTTGC	4681		Db	5537	GTCTTTATGGCATTCGATTTGATTTGATATCAGGAAATCATTTGTCCCATTGGAGTTTC	5596	
Db	4457	GTGATGTCATCGGAAGATTTTAATGAAGAAACAGCTCCGAAACAAAATTCGATTTGC	4516		Qy	5762	CAGAAACACCAAGAAAGCGCTACACCTCAGAGGAAAGCGCTTCGATCAAGTGCACTGC	5821	
Qy	4682	TGAGTCTCTCAGATGCTGAAGTTAACTTACCGAGATAGCCTTTGAGACCTGCCATCAACCA	4741		Db	5597	CAGAAACACCAAGAAAGCGCTTACACCTCAGAGGAAAGCGCTTCGATCAAGTGCACTGC	5856	
Db	4517	TGAGTCTCTCAGATGCTGAAGTTAACTTACCGAGATAGCCTTTGAGACCTGCCATCAACCA	4576		Qy	5822	GGCAAGAGACACAGAAAGCGCCCAAGCAAACTGGCCCTGTATTATTGAAACCTGGGTAG	5881	
Qy	4742	AAGAGTCTGACAGTACACAGACGACACACCCCTCAGCATCTTGTCCAGAAAGCAATTCAG	4801		Db	5657	GGCAAGAGACACAGAAAGCGCCCAAGCAAACTGGCCCTGTATTATTGAAACCTGGGTAG	5716	
Db	4577	AAGAGTCTGACAGTACACAGACGACACCCCTCAGCATCTTGTCCAGAAAGCAATTCAG	4636		Qy	5882	CAGAAGAAGAACTGGAAATTTGTGGGAGATCAGGCGATTTGCTGAGAGAGTGGAGAAAGAA	5941	
Qy	4802	TTAATCAGTGAAGATPATGGAAATAGAAACCTCAGAAAGTTAAGAAAGTTACTTCAATCAC	4861		Db	5717	CAGAAGAAGAACTGGAAATTTGTGGGAGATCAGGCGCATTTGCTGAGAGAGTGGAGAAAGAA	5776	
Db	4637	TTAATCAGTGAAGATPATGGAAATAGAAACCTCAGAAAGTTAAGAAAGTTACTTCAATCAC	4696		Qy	5942	AGSCAACAGCAGTTGAGCAACAGCGCTTAAGAAACGACTGGAGCAGCAGAGCCACAGTGA	6001	
Qy	4862	CTATTACTCTGAGAGGAATCTAATCTCAGTAATGACTTTATTGTATGAAATGGTCTGC	4921		Db	5777	AGSCAACAGCAGTTGAGCAACAGCGCTTAAGAAACGACTGGAGCAGCAGAGCCACAGTGA	5836	
Db	4697	CTATTACTCTGAGAGGAATCTAATCTCAGTAATGACTTTATTGTATGAAATGGTCTGC	4756		Qy	6002	TTGCAACTTCCACTACTCTCCCAACAGCAGTACACACAGCAGCAGCATCTCTCCAGCAGAGA	6061	
Qy	4922	CCATCAACAAAATGAAAATGTCAATGGAGAACTTAAAGAAAAACCGTCATCACAGAAG	4981		Db	5837	TTGCAACTTCCACTACTCTCCCAACAGCAGTACACACAGCAGCAGCATCTCTCCAGCAGAGA	5896	
Db	4757	CCATCAACAAAATGAAAATGTCAATGGAGAACTTAAAGAAAAACCGTCATCACAGAAG	4816		Qy	6062	AGTTATGTGGGCCCTTAAGTGGCTCAGTTTACAATCGGAACCAAAATGGTACTAATCTA	6121	
Qy	4982	TCACCGAGTACCTCCACAGTGCCACAGAAATCAAAAATCTGATCAAGGTAGAAAAAG	5041		Db	5897	AGTTATGTGGGCCCTTAAGTGGCTCAGTTTACAATCGGAACCAAAATGGTACTAATCTA	5956	
Db	4817	TCACCGAGTACCTCCACAGTGCCACAGAAATCAAAAATCTGATCAAGGTAGAAAAAG	4876		Qy	6122	CTAAGTTGAGTCTCCAGCTTACAGTAACTTCAACAAACAAACAGAACTTTTCATCAAACT	6181	
Qy	5042	CGGATAGCAAACTGGTTTCTTCCACAGAAATTTGCAAAATCCACTGTCCACACCA	5101		Db	5957	CTAAGTTGAGTCTCCAGCTTACAGTAACTTCCACAAACAAACAGAACTTTTCATCAAACT	6016	
Db	4877	CGGATAGCAAACTGGTTTCTTCCACAGAAATTTGCAAAATCCACTGTCCACACCA	4936		Qy	6182	TTGCTACATGGGTTAAGCAGGCGCAGTCAAAATTCAGGCGTGTGTTCAAGTACAGCAGAAAG	6241	
Qy	5102	CCACTACAACAGTGACCAAGCTTTCCACACCCCTCCACAGCGGCGAGTGTGGACATCATCT	5161		Db	6017	TTGCTACATGGGTTAAGCAGGCGCAGTCAAAATTCAGGCGTGTGTTCAAGTACAGCAGAAAG	6076	
Db	4937	CCACTACAACAGTGACCAAGCTTTCCACACCCCTCCACAGCGGCGAGTGTGGACATCATCT	4996						

Db	6257	TCCAAAGTCAACACTAGGAAAGCAATATTTCGAACACCTGTGATGGTACAGCCAGGTG	6316
Qy	6482	CTCCTCAGCAAGTGAATGACTCAAAATCATCAGGGGGAGAGCTGTCTCCACTGCAAGTCTCCG	6541
Db	6317	CTCCTCAGCAAGTGAATGACTCAAAATCATCAGGGGGAGAGCTGTCTCCACTGCAAGTCTCCG	6376
Qy	6542	CCCCTAACACGGTTTCTCTCAACACCTGGGCAGAAAAGCTTAACTTCAGCAACCTGCCACTT	6601
Db	6377	CCCCTAACACGGTTTCTCTCAACACCTGGGCAGAAAAGCTTAACTTCAGCAACCTGCCACTT	6436
Qy	6602	CAAAATACAGTCTTTCAGGCTCAACACCCCTCGCCCCAACAAAGGACAAAGTGAAGCTCA	6661
Db	6437	CAAAATACAGTCTTTCAGGCTCAACACCCCTCGCCCCAACAAAGGACAAAGTGAAGCTCA	6496
Qy	6662	CCATGGCTCAACTTACTCAGTTAAACAGAGGCCAGGTGGCAATCAAGGTTTGACAGTAG	6721
Db	6497	CCATGGCTCAACTTACTCAGTTAAACAGAGGCCAGGTGGCAATCAAGGTTTGACAGTAG	6556
Qy	6722	TAATTCAGGACAAGGTCAAACTACTGGACAGTTTGCAAGTTGATACCTCAAGGGGTGACTG	6781
Db	6557	TAATTCAGGACAAGGTCAAACTACTGGACAGTTTGCAAGTTGATACCTCAAGGGGTGACTG	6616
Qy	6782	TACTCCAGGCCAGGCCAGAGCTAATGCAAGCTGCAATGCCAAATGGTACTGTTCCAGC	6841
Db	6617	TACTCCAGGCCAGGCCAGAGCTAATGCAAGCTGCAATGCCAAATGGTACTGTTCCAGC	6676
Qy	6842	GATTCTCTTTTACCCCATTTGGCAACACAGCCACACAGCCAGCCACCAACACCACTG	6901
Db	6677	GATTCTCTTTTACCCCATTTGGCAACACAGCCACACAGCCAGCCACCAACCACTG	6736
Qy	6902	TTTCCACGACAGCAGCGGTACAGGTGAACAAAGGCGAGAGTAAACTGTCAACCCAGATGC	6961
Db	6737	TTTCCACGACAGCAGCGGTACAGGTGAACAAAGGCGAGAGTAAACTGTCAACCCAGATGC	6796
Qy	6962	AGGTACATCAAGACAAACCCCTGCCACGAGTCAAGTCAATCAAGTGGGTCCACCAAG	7021
Db	6797	AGGTACATCAAGACAAACCCCTGCCACGAGTCAAGTCAATCAAGTGGGTCCACCAAG	6856
Qy	7022	CCAGCCACAGACTGCTCAGCGCTTCAGCTCGGCCCGCCAGCCCAACCCAGCCAGTCCC	7081
Db	6857	CCAGCCACAGACTGCTCAGCGCTTCAGCTCGGCCCGCCAGCCCAACCCAGCCAGTCCC	6916
Qy	7082	CAGCTCAGCGCTGAAGTTCAGACTCAGCTGAAGTTCAGACCCCAACAACTGTTTCATCCC	7141
Db	6917	CAGCTCAGCGCTGAAGTTCAGACTCAGCTGAAGTTCAGACCCCAACAACTGTTTCATCCC	6976
Qy	7142	ATGTCCCTTCTGAAGCAACACCCACGACAGTCAATCAAGCCCAAGTTGCAGCAC	7201
Db	6977	ATGTCCCTTCTGAAGCAACACCCACGACAGTCAATCAAGCCCAAGTTGCAGCAC	7036
Qy	7202	AGTCTCAGCGCTCAAAAGTAAATGTCCAAGGACAGTCTCCTGTTGCTGTCCAAGTCCATCAC	7261
Db	7037	AGTCTCAGCGCTCAAAAGTAAATGTCCAAGGACAGTCTCCTGTTGCTGTCCAAGTCCATCAC	7096
Qy	7262	AGACTCGAATAGTCCATCAACTCCACTCCCACTGTCTCTGGACAAACATCCCAAGTTC	7321
Db	7097	AGACTCGAATAGTCCATCAACTCCACTCCCACTGTCTCTGGACAAACATCCCAAGTTC	7156
Qy	7322	AGACTACAACTCAACCGGATTCCAATTCAACACATACATCTCTTCAGATACCTTCCC	7381
Db	7157	AGACTACAACTCAACCGGATTCCAATTCAACACATACATCTCTTCAGATACCTTCCC	7216
Qy	7382	AAGGCCAGCCAGTCAACACCCAGGTACAGTCTTCAACTCAAACTCTTTCATCAGGAC	7441
Db	7217	AAGGCCAGCCAGTCAACACCCAGGTACAGTCTTCAACTCAAACTCTTTCATCAGGAC	7276
Qy	7442	AAACTTTAATCAAGTTAGTGTTCATCCCATCCGTCCTCAGCTACAAATACAGCAGC	7501
Db	7277	AAACTTTAATCAAGTTAGTGTTCATCCCATCCGTCCTCAGCTACAAATACAGCAGC	7336
Qy	7502	CACAGCCCAAGTCAATGCTGTGCCTCAGCTGCAACAAAGTCCAGGTTCCTCTCAGA	7561
Db	7337	CACAGCCCAAGTCAATGCTGTGCCTCAGCTGCAACAAAGTCCAGGTTCCTCTCAGA	7396
Qy	7562	TCCAGTACAGGTTGTGGCTCAGATACAGGCTCAGCAAAAGTGGTGTGCCCCAGCAAAATCA	7621
Db	7397	TCCAGTACAGGTTGTGGCTCAGATACAGGCTCAGCAAAAGTGGTGTGCCCCAGCAAAATCA	7456
Qy	7622	AACTCCAGTTACCTATCCAAATTCAGCAAAAGCAGTCTGTGCAGACTCACAGATTCAGA	7681
Db	7457	AACTCCAGTTACCTATCCAAATTCAGCAAAAGCAGTCTGTGCAGACTCACAGATTCAGA	7516
Qy	7682	ATGTGGTTACAGTGCAGGAGCCAGTGTCAAGAGCAGTTGCAAAAGGGTTCAGCAACTCA	7741
Db	7517	ATGTGGTTACAGTGCAGGAGCCAGTGTCAAGAGCAGTTGCAAAAGGGTTCAGCAACTCA	7576
Qy	7742	GGGATCAGCAGCAAAAAGAAAGAAACAGCAACAGATAGAAATTAACCTGAAACACCCCTCC	7800
Db	7577	GGGATCAGCAGCAAAAAGAAAGAAACAGCAACAGATAGAAATTAACCTGAAACACCCCTCC	7636
Qy	7801	AACTTCTTAATCAAAAGTTGAAATCAATCAGAAACAGTGGTGTATGAAGCATATGCTGTA	7860
Db	7637	AACTTCTTAATCAAAAG-TGAAATCAATCAGAAACAGTGGTGTATGAAGCATATGCTGTA	7695
Qy	7861	ATAGAACTTTAAACAGAAAAAGAGCATGACTCCAGCTGAAAGAGAGAGATCAAAAGA	7920
Db	7696	ATAGAACTTTAAACAGAAAAAGAGCATGACTCCAGCTGAAAGAGAGAGATCAAAAGA	7755
Qy	7921	ATGATTGTCTGTAAACAGAGTGAAGTATATTTTGGATGAAGATAGATTAAGAAAGAAAA	7980
Db	7756	ATGATTGTCTGTAAACAGAGTGAAGTATATTTTGGATGAAGATAGATTAAGAAAGAAAA	7815
Qy	7981	CAGGCAGCAAAAACCGAAGCGTGAAGAGAGTGTGGAGCAGAAACCTAGCAAGCAGAAAT	8040
Db	7816	CAGGCAGCAAAAACCGAAGCGTGAAGAGAGTGTGGAGCAGAAACCTAGCAAGCAGAAAT	7875
Qy	8041	GCCACTAAGCTGTGAGTCTGCTCTTCAAGCAAAAAGAGCAGCTCAGAGCCGAGATGCTG	8100
Db	7876	GCCACTAAGCTGTGAGTCTGCTCTTCAAGCAAAAAGAGCAGCTCAGAGCCGAGATGCTG	7935
Qy	8101	AAGAAGAGAGCACTCTCGGACAGGATCTGCAAAATGAAGTGCAGGAGAGCTGGAAGAGA	8160
Db	7936	AAGAAGAGAGCACTCTCGGACAGGATCTGCAAAATGAAGTGCAGGAGAGCTGGAAGAGA	7995
Qy	8161	GACCTGAAATTAAGAAAGAAAAAGACCTGATGCAGTTGGCTCAGGCCACAGCAGTAGCT	8220
Db	7996	GACCTGAAATTAAGAAAGAAAAAGACCTGATGCAGTTGGCTCAGGCCACAGCAGTAGCT	8055
Qy	8221	GCACCTTGCCTCCAGTGCACACAGTCTTCCAGCCCTCCAGCCCTCCACCTTCCACT	8280
Db	8056	GCACCTTGCCTCCAGTGCACACAGTCTTCCAGCCCTCCAGCCCTCCACCTTCCACT	8115
Qy	8281	CCCCCTCCAGTGTGTGCAACACACAGGCTTCTGTCCACGCCACCTTACCTTGTGCT	8340
Db	8116	CCCCCTCCAGTGTGTGCAACACACAGGCTTCTGTCCACGCCACCTTACCTTGTGCT	8175
Qy	8341	TCCCAGAGAGGAAAGCGGAGAGGAAAAAGACTCCAGCTCAAAAGTCCAGAAAAAGAAA	8400
Db	8176	TCCCAGAGAGGAAAGCGGAGAGGAAAAAGACTCCAGCTCAAAAGTCCAGAAAAAGAAA	8235
Qy	8401	ATGATCTCTACTACTCTCAAGGAAACTTAAGNAGGACACAAAGCTTTACTGTATCTGTA	8460
Db	8236	ATGATCTCTACTACTCTCAAGGAAACTTAAGNAGGACACAAAGCTTTACTGTATCTGTA	8295
Qy	8461	ACGCTTATGATGAATCTAAATTTTATATTTGGCTGTGATCGGTGTGAGATTTGTTACCAT	8520
Db	8296	ACGCTTATGATGAATCTAAATTTTATATTTGGCTGTGATCGGTGTGAGATTTGTTACCAT	8355
Qy	8521	GGGCGCTGCTGTGGCATCTTGCAAAGTGAGGAGAGCTCATTTGATGAGTATGCTGTCCA	8580
Db	8356	GGGCGCTGCTGTGGCATCTTGCAAAGTGAGGAGAGCTCATTTGATGAGTATGCTGTCCA	8415
Qy	8581	CAGTCCCAAGTCAACAGAGGATGCCATGACGTGCTCGGCCACTTAAACAGAGAGGATAT	8640
Db	8416	CAGTCCCAAGTCAACAGAGGATGCCATGACGTGCTCGGCCACTTAAACAGAGAGGATAT	8475

QY	2299	GAAATTTACGGTTGAGCACCACCAAAAGGAAGTGCATCATGAAAGGAAATATCAACAATTTAT	2358	3379	AAAAAAAATCATCCAACTAGATGCATCTCTTTGAAAGGAGAATTAACACAGTTTACACTG	3438
Db	1852	GAAATTTACGGTTGAGCACCACCAAAAGGAAGTGCATCATGAAAGGAAATATCAACAATTTAT	1911	2932	AAAAAAAATCATCCAACTAGATGCATCTCTTTGAAAGGAGAATTAACACAGTTTACACTG	2991
QY	2359	TTTAAATTTGGTCAACAAGGGAAGTATCGGCTCTACCAACAATCAATACATCCACCAATTTCA	2418	3439	GAAGAAAAACAGCGACTCGAAAAAATCAAGTTGAGGGTGGAAATTAAGGGTATAGGAAG	3498
Db	1912	TTTAAATTTGGTCAACAAGGGAAGTATCGGCTCTACCAACAATCAATACATCCACCAATTTCA	1971	2992	GAAGAAAAACAGCGACTCGAAAAAATCAAGTTGAGGGTGGAAATTAAGGGTATAGGAAG	3051
QY	2419	TTTGGCTTTGAATAAGCACCAGCACAGAGAGACCATGATAAGAGAAGGCATCTTGCACAT	2478	3499	ACTTCTACAAATTTCTCAAAAAATCTCTGTAATCACCACTACCAAGTAAATACGAAAGCAAAAGAA	3558
Db	1972	TTTGGCTTTGAATAAGCACCAGCACAGAGAGACCATGATAAGAGAAGGCATCTTGCACAT	2031	3052	ACTTCTACAAATTTCTCAAAAAATCTCTGTAATCACCACTACCAAGTAAATACGAAAGCAAAAGAA	3111
QY	2479	AAGTCTCTGCTGACTCCAGCAGAGAGTTCAAAATGGAACGGTTCTGTCTCATGGGTCCAAA	2538	3559	GGGTGTACAGTGCATCGATGAGACAAGAACAGAGCCCAAAATGCAAAATATATGATCAACCT	3618
Db	2032	AAGTCTCTGCTGACTCCAGCAGAGAGTTCAAAATGGAACGGTTCTGTCTCATGGGTCCAAA	2091	3112	GGGTGTACAGTGCATCGATGAGACAAGAACAGAGCCCAAAATGCAAAATATATGATCAACCT	3171
QY	2539	GTTCCTTACCATACTACTCTGAGACTGACTATCCACCAATATAGAAAACAACATCCCTTTCA	2598	3619	GAGGACTTGATTTCAGGGATGTTCCACAAAGTGATTCTCAGTCTCTTGAATAGATGATGATCCT	3678
Db	2092	GTTCCTTACCATACTACTCTGAGACTGACTATCCACCAATATAGAAAACAACATCCCTTTCA	2151	3172	GAGGACTTGATTTCAGGGATGTTCCAGAAAGTGATTCTCAGTCTCTTGAATAGATGATGATCCT	3231
QY	2599	TCCTTTTTCATCCCACTGGGCATCACATAGGGCAAAATTTGGATCAAGCAGTTCAGATG	2658	3679	AGTCATACCACAAAACAACTTTATCCAAAAGATFCGAGTGTTAGATGATGTCTCCATTCGG	3738
Db	2152	TCCTTTTTCATCCCACTGGGCATCACATAGGGCAAAATTTGGATCAAGCAGTTCAGATG	2211	3232	AGTCATACCACAAAACAACTTTATCCAAAAGATFCGAGTGTTAGATGATGTCTCCATTCGG	3291
QY	2659	TGTAGCAAAACCCAGAGAAATTTGCATTTGGCTTTAGCCATTTTGGAGTGTGCAGTTAAACCA	2718	3739	AGCCAGAAAACAAAATTCGCGAAACAAAATTCATTGAAATGACATAGAGAAGAAAGTC	3798
Db	2212	TGTAGCAAAACCCAGAGAAATTTGCATTTGGCTTTAGCCATTTTGGAGTGTGCAGTTAAACCA	2271	3292	AGCCAGAAAACAAAATTCGCGAAACAAAATTCATTGAAATGACATAGAGAAGAAAGTC	3351
QY	2719	GTTCGTATGCTACCAATATGGCGAGAAATTTTAGGACATACCAGTTACACCGGATGACA	2778	3799	TCGTGACTTGGCCAGTAGAGCCAGAACCCACTTAAGAGTAAACACCAAGGAAATGATTTT	3858
Db	2272	GTTCGTATGCTACCAATATGGCGAGAAATTTTAGGACATACCAGTTACACCGGATGACA	2331	3352	TCGTGACTTGGCCAGTAGAGCCAGAACCCACTTAAGAGTAAACACCAAGGAAATGATTTT	3411
QY	2779	TCAATTTGAAGAGAGAGAGAGAAAGTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG	2838	3859	TTTCATCGATGACTCTAACTAGCCAGTGCAGATGATATTGGTACTTTGATCTCTTAAGAAC	3918
Db	2332	TCAATTTGAAGAGAGAGAGAGAGAGTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG	2391	3412	TTTCATCGATGACTCTAACTAGCCAGTGCAGATGATATTGGTACTTTGATCTCTTAAGAAC	3471
QY	2839	GAACGATGCAGAACGACATGGGTAAATACACATTTCCAGTTTAAGCATCAGGTTTGG	2898	3919	AAAAACCGCTCATACAGGAGGAAAGTGACACCATTTGTTCTTCTTCCAAAGATGCTTTA	3978
Db	2392	GAACGATGCAGAACGACATGGGTAAATACACATTTCCAGTTTAAGCATCAGGTTTGG	2451	3472	AAAAACCGCTCATACAGGAGGAAAGTGACACCATTTGTTCTTCTTCCAAAGATGCTTTA	3531
QY	2899	AAACAAAAAGGTGAAGAGTACAGAGTACAGAGATATGGTGGTGGAGCTGGATTTAGTAA	2958	3979	CATTTCATGCTGCTAAAAAGTACCAATGACAGAGATGCCACACTCTGTCAAGAGCAATG	4038
Db	2452	AAACAAAAAGGTGAAGAGTACAGAGTACAGAGATATGGTGGTGGAGCTGGATTTAGTAA	2511	3532	CATTTCATGCTGCTAAAAAGTACCAATGACAGAGATGCCACACTCTGTCAAGAGCAATG	3591
QY	2959	ACTCATGTTTATAGTTTCTCCTAAATTCAGGCAATACTAATGTGAATTTACAGAAAG	3018	4039	GACTTTGAAAGGAAAACTGGGATGTGACTCTGAAATCTAATAGCACTTTTGGAAAAATAGTTCT	4098
Db	2512	ACTCATGTTTATAGTTTCTCCTAAATTCAGGCAATACTAATGTGAATTTACAGAAAG	2571	3592	GACTTTGAAAGGAAAACTGGGATGTGACTCTGAAATCTAATAGCACTTTTGGAAAAATAGTTCT	3651
QY	3019	TGCTTAGAAGGAACCAAAAAATATATGGATGAAATATGATGATGATGATGATGATGATGAT	3078	4099	GATACCGTGTCTATTCCAGGATAGCAGTGAAGAGATATGATTTTCCAGAAATAGCAATGAA	4158
Db	2572	TGCTTAGAAGGAACCAAAAAATATATGGATGAAATATGATGATGATGATGATGATGATGAT	2631	3652	GATACCGTGTCTATTCCAGGATAGCAGTGAAGAGATATGATTTTCCAGAAATAGCAATGAA	3711
QY	3079	TGTTACAGAGTCCAAAAAATAAATAAGAGGCTGATCTCAAAAAAGATGAGGTAA	3138	4159	AGCATTTCGAAACAGTTTCAGAACTCGAGAACTGTGAGGACAGGCTGCGGGTCAAGGGG	4218
Db	2632	TGTTACAGAGTCCAAAAAATAAATAAGAGGCTGATCTCAAAAAAGATGAGGTAA	2691	3712	AGCATTTCGAAACAGTTTCAGAACTCGAGAACTGTGAGGACAGGCTGCGGGTCAAGGGG	3771
QY	3139	GGTTTCAGATGCTGCAAAAGGAGCAGACCAAAATGAAATGGAATGATCTCAAAAGATTTACTGAG	3198	4219	TGTGAGTTGGTTTCTGGTGGTCCACTGGAAACTGTGAGGACAGGCTGCGGGTCAAGGGG	4278
Db	2692	GGTTTCAGATGCTGCAAAAGGAGCAGACCAAAATGAAATGGAATGATCTCAAAAGATTTACTGAG	2751	3772	TGTGAGTTGGTTTCTGGTGGTCCACTGGAAACTGTGAGGACAGGCTGCGGGTCAAGGGG	3831
QY	3199	AGAAGGACCAAGATGTGAAGGAGCTCTTTAGATCTGACAGTGATAAACCCCTGCAAGGAA	3258	4279	ACTGAAGCAAAATGGTAAAAAACCAGTACGAGAGAAATTTAGAGAGAGAGACCAAGTTAAT	4338
Db	2752	AGAAGGACCAAGATGTGAAGGAGCTCTTTAGATCTGACAGTGATAAACCCCTGCAAGGAA	2811	3832	ACTGAAGCAAAATGGTAAAAAACCAGTACGAGAGAAATTTAGAGAGAGAGACCAAGTTAAT	3891
QY	3259	GAACCAATGGAATGACATGACATGAAAAACAGAGTACATGTAATTTGTCAGGAGAT	3318	4339	AAATGCTAGTGAATCAATTAAGCTTAAAAATACCACCTGACAAAAAGAAATTAATGAAAAATCGA	4398
Db	2812	GAACCAATGGAATGACATGACATGAAAAACAGAGTACATGTAATTTGTCAGGAGAT	2871	3892	AAATGCTAGTGAATCAATTAAGCTTAAAAATACCACCTGACAAAAAGAAATTAATGAAAAATCGA	3951
QY	3319	TCTCAAGTATGATGTTGATCAATGTTAGTGAGGGTTTTTCATCTAAGGACTAGTTTACAAAAAG	3378	4399	GAGTCTGAAAAAGAAAGGACAGAGCAAGTACATTTCAAAATTAATGAAAAAGATTAATAA	4458
Db	2872	TCTCAAGTATGATGTTGATCAATGTTAGTGAGGGTTTTTCATCTAAGGACTAGTTTACAAAAAG	2931	3952	GAGTCTGAAAAAGAAAGGACAGAGCAAGTACATTTCAAAATTAATGAAAAAGATTAATAA	4011
				4459	CCCAAAATATATTTGAAAGGTGAATGCTTTGAAAGAAATTTTCTGAGAGTAGAGTAGTAAGT	4518

|||||
4012 CCCAAATATATTGAAAGGTGAATGCTTGAAGAAATTTCTGAGAGTAGTAGTAAAGT 4071 Db
QY
4519 GGTAAATGTTGAACCAAGGTTAAATAATATAATAATAAATCCCTGAGAAATGATATATAA 4578 QY
Db
4072 GGTAAATGTTGAACCAAGGTTAAATAATATAATAATAAATCCCTGAGAAATGATATATAA 4131 Db
QY
4579 TCATTGACTGTTAAAGAAATCTGCTATAAAGCCATTCAATTAATGGTGATGTCATCATGAA 4638 QY
Db
4132 TCATTGACTGTTAAAGAAATCTGCTATAAAGCCATTCAATTAATGGTGATGTCATCATGAA 4191 Db
QY
4639 GATTTTAAATGAAGAAACAGCTCCGAAACAAATCGCATTTGCTGAGTCTTCTCAGATGCT 4698 QY
Db
4192 GATTTTAAATGAAGAAACAGCTCCGAAACAAATCGCATTTGCTGAGTCTTCTCAGATGCT 4251 Db
QY
4699 GAAGGTAACTACCGAGATAGCCTTTGAGACCCCTGCCATCAACCAAGAGTCTGACAGTACA 4758 QY
Db
4252 GAAGGTAACTACCGAGATAGCCTTTGAGACCCCTGCCATCAACCAAGAGTCTGACAGTACA 4311 Db
QY
4759 CAGAGGACCAACCCCTCAGCATCTTGTCCAGAAAGCAATTCAGTTAATCAGGTAGAAAGAT 4818 QY
Db
4312 CAGAGGACCAACCCCTCAGCATCTTGTCCAGAAAGCAATTCAGTTAATCAGGTAGAAAGAT 4371 Db
QY
4819 ATGGAATATGAACCTCAGAAAGTTAAGAAAGTTACTTCATCACCTATTACTTCGTAAGAG 4878 QY
Db
4372 ATGGAATATGAACCTCAGAAAGTTAAGAAAGTTACTTCATCACCTATTACTTCGTAAGAG 4431 Db
QY
4879 GAATCTAATCTCAGTAATGACTTTTATGATGAATGGTCTGCCCATCAACAAAAATGAA 4938 QY
Db
4432 GAATCTAATCTCAGTAATGACTTTTATGATGAATGGTCTGCCCATCAACAAAAATGAA 4491 Db
QY
4939 AATGTCATGGAGATCTTAAAGAAACCCCTCATCAGAAAGTCAACCAAGATGACCTCC 4998 QY
Db
4492 AATGTCATGGAGATCTTAAAGAAACCCCTCATCAGAAAGTCAACCAAGATGACCTCC 4551 Db
QY
4999 ACAGTGGCCACAGAAATCAAAATCTGTATCAAGGTAGAAAGGGGATAAGCAACACTGTG 5058 QY
Db
4552 ACAGTGGCCACAGAAATCAAAATCTGTATCAAGGTAGAAAGGGGATAAGCAACACTGTG 4611 Db
QY
5059 GTTTCCTCCACAGAAATTTGTCAAAATCCACTGTCAACACCAACCACTTACAAACAGTGACC 5118 QY
Db
4612 GTTTCCTCCACAGAAATTTGTCAAAATCCACTGTCAACACCAACCACTTACAAACAGTGACC 4671 Db
QY
5119 AAGCTTTCCACACCTCCACAGGGGCGAGTGTGGACATCATCTCTGTAAAGAGGACAGAC 5178 QY
Db
4672 AAGCTTTCCACACCTCCACAGGGGCGAGTGTGGACATCATCTCTGTAAAGAGGACAGAC 4731 Db
QY
5179 AAAACCGTGGTCCACACGACAGTGACAGACTCCCTGACCAACCGGAGGACACACTGTT 5238 QY
Db
4732 AAAACCGTGGTCCACACGACAGTGACAGACTCCCTGACCAACCGGAGGACACACTGTT 4791 Db
QY
5239 ACATCTATGACTGTGAGCAAAAGAGTATTCACACAGAGACAAAGTGAATGATAATTT 5298 QY
Db
4792 ACATCTATGACTGTGAGCAAAAGAGTATTCACACAGAGACAAAGTGAATGATAATTT 4851 Db
QY
5299 TCAAGACCAAGAGAGAGTGTGAGTACAGTCTGACATCTGACATCTGATAGAAATTTGTACC 5358 QY
Db
4852 TCAAGACCAAGAGAGAGTGTGAGTACAGTCTGACATCTGACATCTGATAGAAATTTGTACC 4911 Db
QY
5359 AAGAGCACCAGAGAGAGATTTTGTGCTTAATGATGACTTAAAAAAGTTGGCCCGA 5418 QY
Db
4912 AAGAGCACCAGAGAGAGATTTTGTGCTTAATGATGACTTAAAAAAGTTGGCCCGA 4971 Db
QY
5419 AAAGGAGGAATCCGAGAGGTCCCTTATTTTAAATACAATGCAAAACCTGCTTTGGATATA 5478 QY
Db
4972 AAAGGAGGAATCCGAGAGGTCCCTTATTTTAAATACAATGCAAAACCTGCTTTGGATATA 5031 Db
QY
5479 TGGCCATATCTTCTCAGACCGACCTTTGACATCACTTGAGGTATAGACTTCAGACA 5538 QY
Db
5032 TGGCCATATCTTCTCAGACCGACCTTTGACATCACTTGAGGTATAGACTTCAGACA 5091 Db
QY
5539 GTAAAGTCTTACGTGAGTACGCTGATGTTACGGTACTCTGGGCAAGTTTGAGATGG 5598 QY
|||||

5092 GTAAAGTCTTACGTGAGCTGAGCTGATGTTACGGTACTCTGGGCAAGTTTGAGATGG 5151 Db
QY
5599 GATGATATGCGGCCCAAGGTTCTCCAGCAGGAGGAGTACACGACAGAAAAATCCCGAA 5658 Db
5152 GATGATATGCGGCCCAAGGTTCTCCAGCAGGAGGAGTACACGACAGAAAAATCCCGAA 5211 Db
QY
5659 ACTGAAATCACAAACAGAAATTAATTAAGAGGAGAGATGTTGGTCTTATGGCATTCGA 5718 Db
5212 ACTGAAATCACAAACAGAAATTAATTAAGAGGAGAGATGTTGGTCTTATGGCATTCGA 5271 Db
QY
5719 TTTGAAATATGATCAGGAAAAATCATTTGTCATTTGGAGTTCCAGAAAAACCAAAAGAA 5778 Db
5272 TCTGAAATATGATCAGGAAAAATCATTTGTCATTTGGAGTTCCAGAAAAACCAAAAGAA 5331 Db
QY
5779 ACSCCTACACCTCAGAGGAAAGCCCTTCGATCAAGTGCACCTGGCGCAAGAGACAGAA 5838 Db
5332 ACSCCTACACCTCAGAGGAAAGCCCTTCGATCAAGTGCACCTGGCGCAAGAGACAGAA 5391 Db
QY
5839 ACSCCAAGCAAACTGGCCCTGTTTATTTATTTGAACCTGGGTAGCAGAAAGAACTGGAA 5898 Db
5392 ACSCCAAGCAAACTGGCCCTGTTTATTTATTTGAACCTGGGTAGCAGAAAGAACTGGAA 5451 Db
QY
5899 TTTGGSAGATCAGGCACTTTGCTGAGAGAGTGGAGAAAGAAAGCAAGCAGTTGAG 5958 Db
5452 TTTGGSAGATCAGGCACTTTGCTGAGAGAGTGGAGAAAGAAAGCAAGCAGTTGAG 5511 Db
QY
5959 CAACAGGCTAAGAAACGACTGGAGCAGCAGAGCCGACAGTGAATGCAACTTCCACTACT 6018 Db
5512 CAACAGGCTAAGAAACGACTGGAGCAGCAGAGCCGACAGTGAATGCAACTTCCACTACT 5571 Db
QY
6019 TCCCAACAGCAGTACAACACGACCACTCTCCAGCAGAGAAAGTTATGTTGGCCCC 6078 Db
5572 TCCCAACAGCAGTACAACACGACCACTCTCCAGCAGAGAAAGTTATGTTGGCCCC 5631 Db
QY
6079 ATAAGTGGCTCAGTTACAACCTGGAACCAAAATGGTACTACTACTAAAGTTGGATTCGA 6138 Db
5632 ATAAGTGGCTCAGTTACAACCTGGAACCAAAATGGTACTACTACTAAAGTTGGATTCGA 5691 Db
QY
6139 GCTACAGTACATTTCCAAACAAACAGAACTTTCACTCAAACTTTGCTACATGGGTTAAG 6198 Db
5692 GCTACAGTACATTTCCAAACAAACAGAACTTTCACTCAAACTTTGCTACATGGGTTAAG 5751 Db
QY
6199 CAAGCCAGTCAAAATTCAGGCTGTTTCAAGTACAGCAGAAAGTCTCTGGTATCATTCGA 6258 Db
5752 CAAGCCAGTCAAAATTCAGGCTGTTTCAAGTACAGCAGAAAGTCTCTGGTATCATTCGA 5811 Db
QY
6259 TCAAGTACAGGTACAGTACAGCAAACTTTACTTTCTATTTCCAGCCAGGACAGCAACAGTC 6318 Db
5812 TCAAGTACAGGTACAGTACAGCAAACTTTACTTTCTATTTCCAGCCAGGACAGCAACAGTC 5871 Db
QY
6319 ACAATTAGGCCCAATACCTCAGGCTCTGGAGGACCAACAGCAATTCACAAGTAATCACA 6378 Db
5872 ACAATTAGGCCCAATACCTCAGGCTCTGGAGGACCAACAGCAATTCACAAGTAATCACA 5931 Db
QY
6379 GGGCTCAGATTCGCCCTGTTATGACCGTGTATGAGAACACCACTTCCAAACAGTCAACACTA 6438 Db
5932 GGGCTCAGATTCGCCCTGTTATGACCGTGTATGAGAACACCACTTCCAAACAGTCAACACTA 5991 Db
QY
6439 GGAAGGCAATTTATTCGAACACCTGTTGATGGTACAGCCAGGTGCTCTCAGCAAGTGTG 6498 Db
5992 GGAAGGCAATTTATTCGAACACCTGTTGATGGTACAGCCAGGTGCTCTCAGCAAGTGTG 6051 Db
QY
6499 ACTCAATATCATCAGGGGACGCTCTCCACTCAGTCTCCGCCCTTAACACGGTTTCC 6558 Db
6052 ACTCAATATCATCAGGGGACGCTCTCCACTCAGTCTCCGCCCTTAACACGGTTTCC 6111 Db
QY
6559 TCAACACTTGGCAGAAAAAGCTTAACCTTCAGAACCTGCCACTTCAAAATATACAGTCTTCA 6618 Db
6112 TCAACACTTGGCAGAAAAAGCTTAACCTTCAGAACCTGCCACTTCAAAATATACAGTCTTCA 6171 Db
QY
6619 GCCTCACAACCCCTCGCCCCCAACAGGACAAAGTGAAGTCAACCTGCTCAACTTACT 6678 Db
6172 GCCTCACAACCCCTCGCCCCCAACAGGACAAAGTGAAGTCAACCTGCTCAACTTACT 6231 Db

QY	6679	CAGTTAACACAGGGCCACGGTGGCAATCAAGGTTTGACAGTAGTAATTAAGGACAAGGT	6738	7759	AAGAAACAGCAACAGATAGAAATTAACGTGAACACACACCCCTCCAAGCTTCTTAATCAAGTT	7818
Db	6232	CAGTTAACACAGGGCCACGGTGGCAATCAAGGTTTGACAGTAGTAATTAAGGACAAGGT	6291	6958	-----	6957
QY	6739	CAAACTACTGGACAGTTCGAGTTGATACCTCAAGGGTGACTGTACTCCCAGGCCAGGC	6798	7819	GAATCATTCAGAAACAGGTGCTGATGAAGCATAATGCTGTATAGAACATTTAAACACAG	7878
Db	6292	CAAACTACTGGACAGTTCGAGTTGATACCTCAAGGGTGACTGTACTCCCAGGCCAGGC	6351	6958	-----CAGGTGCTGATGAAGCATAATGCTGTATAGAACATTTAAACACAG	7002
QY	6799	CAGCAGCTAATGCAAGCTGCAATGCCAATGCTACTGTTACGCGATTCTCTTTACCCCA	6858	7879	AAAAAGAGCATGACTCCAGCTGAAAGAGAGAAGATCAAAAGAAATGATTCTGTACCCAG	7938
Db	6352	CAGCAGCTAATGCAAGCTGCAATGCCAATGCTACTGTTACGCGATTCTCTTTACCCCA	6411	7003	AAAAAGAGCATGACTCCAGCTGAAAGAGAGAAGATCAAAAGAAATGATTCTGTACCCAG	7062
QY	6859	TTGGCAACACAGCCACAGCCAGCAGCCACACCCAGCCACACCTGTTTCCACGACAGACGA	6918	7939	GTGATGAAGTATATTTTGGATAGATAGATAAAGAGAAAAACAGCAGCAAAAAACGG	7998
Db	6412	TTGGCAACACAGCCACAGCCAGCAGCCACACCCAGCCACACCTGTTTCCACGACAGACGA	6471	7063	GTGATGAAGTATATTTTGGATAGATAGATAAAGAGAAAAACAGCAGCAAAAAACGG	7122
QY	6919	GCTACAGGTGAACAAGCAGAGTAAGCTGTACCTGAGTGTGAGTGTGAGTGTGAGTGTGAGT	6978	7999	AGCGTGAAGCAGGTGAGCAGAGAAACGTACGACGAGAGTCCACTAAGCTGTCAGCT	8058
Db	6472	GCTACAGGTGAACAAGCAGAGTAAGCTGTACCTGAGTGTGAGTGTGAGTGTGAGTGTGAGT	6531	7123	AAGCGTGAAGCAGGTGAGCAGAGAAACGTACGACGAGAGTCCACTAAGCTGTCAGCT	7182
QY	6979	ACCCCTGCCACAGCTCAGTCAATCAAGTGTGGTCCAGCAAAAGCCAGCCACAGACTGCT	7038	8059	CTGCTCTTCAAGCACAAGAGCAGCTCAGAGCCGAGATCCTGAAAGAGAGAGCACTCCCTG	8118
Db	6532	ACCCCTGCCACAGCTCAGTCAATCAAGTGTGGTCCAGCAAAAGCCAGCCACAGACTGCT	6591	7183	CTGCTCTTCAAGCACAAGAGCAGCTCAGAGCCGAGATCCTGAAAGAGAGAGCACTCCCTG	7242
QY	7039	CAGCCTTCAGCTCGGCCAGCCCAACCCAGCCAGCCAGTCCAGCTCAGCCTGAAAGTT	7098	8119	GACAAAGATCTGCAATTTGAAGTGAGGAAAGAGCTGAAGAGAGACTGAAAAATTAAGAA	8178
Db	6592	CAGCCTTCAGCTCGGCCAGCCCAACCCAGCCAGCCAGTCCAGCTCAGCCTGAAAGTT	6651	7243	GACAAAGATCTGCAATTTGAAGTGAGGAAAGAGCTGAAGAGAGACTGAAAAATTAAGAA	7302
QY	7099	CAGACTCAGCCTGAAGTTTCAGACCCCAACCAACTGTTTCATCCCATGTCCTTCAAGCA	7158	8179	GAAAAGACCTGATGCAGTTGGCTCAGGCCACAGCAGTAGCTGCACCTGCCCCAGTG	8238
Db	6652	CAGACTCAGCCTGAAGTTTCAGACCCCAACCAACTGTTTCATCCCATGTCCTTCAAGCA	6711	7303	GAAAAGACCTGATGCAGTTGGCTCAGGCCACAGCAGTAGCTGCACCTGCCCCAGTG	7362
QY	7159	CAACCCACCCAGCAGTCAATCCAAAGTGTGAGCAAGCTCTCAGCAGCTCAAAAGT	7218	8239	ACACCAAGTTCTTCCAGCCCTCCAGCCCTCCACCTTCACTCCCTCCCTCCCTGCTGTG	8298
Db	6712	CAACCCACCCAGCAGTCAATCCAAAGTGTGAGCAAGCTCTCAGCAGCTCAAAAGT	6771	7363	ACACCAAGTTCTTCCAGCCCTCCAGCCCTCCACCTTCACTCCCTCCCTGCTGTG	7422
QY	7219	AATGTCCAAAGACAGTCTGCTGTTGTCGTTCCAAAGTCCATCAGACGTCGAATAGTCCA	7278	8299	CAACACAGGCTTCTGTCCAGCCCTTACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	8358
Db	6772	AATGTCCAAAGACAGTCTGCTGTTGTCGTTCCAAAGTCCATCAGACGTCGAATAGTCCA	6831	7423	CAACACAGGCTTCTGTCCAGCCCTTACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	7482
QY	7279	TCAACTCCATCCCACTGCTCCTGGACAACAAATCCAGGTTTCAGACTACACCTTCACAA	7338	8359	GAAGAGGAAAAGACTCCAGCTCAAGTCCAAAGAAAAGAAAATGATCTCTACTACTCA	8418
Db	6832	TCAACTCCATCCCACTGCTCCTGGACAACAAATCCAGGTTTCAGACTACACCTTCACAA	6891	7483	GAAGAGGAAAAGACTCCAGCTCAAGTCCAAAGAAAAGAAAATGATCTCTACTACTCA	7542
QY	7339	CGATTTCGAATTCACACCATACATCTCTTCAGATACCTTCCCAAGGGCAGCCAGTCA	7398	8419	AAGGAACTTAAGAGGACACAAAGCTTTACTGTATCTGTATAAACGCTTTATGATGAATCT	8478
Db	6892	CGATTTCGAATTCACACCATACATCTCTTCAGATACCTTCCCAAGGGCAGCCAGTCA	6951	7543	AAGGAACTTAAGAGGACACAAAGCTTTACTGTATCTGTATAAACGCTTTATGATGAATCT	7602
QY	7399	CAACCCAGGTCAGCTTTCAACTCAACTCTTTTCATCAGGACAAACTTTAAATCAAGTT	7458	8479	A-----	8479
Db	6952	CAACCC-----	6957	7603	AAGTTCTATATTGGCTGTGATCTTTGTACTAACTGGTATCATGGAGAAATGTGTGGCATC	7662
QY	7459	AGTGTTCATCCCATCCGTCCTCAGCTACAAATACAGCAGCCACAGCCCAAGTCATT	7518	8480	-----	8479
Db	6958	-----	6957	7663	ACAGAAAAGGAGGCTTAAGAAAATGGATGTGTACATCTGTAATGATTTAAACGGGCACAA	7722
QY	7519	GCTGTGCTCAGTCGAACAACAAGTCCAGGTTCTCTCTCAGATCCAGTTCAGGTTGTG	7578	8480	-----AATTT	8484
Db	6958	-----	6957	7723	GAGGCAGCAGTGAGGAATTTGACTGTATCTCAGAAACACCTTATGATGAGTCACAAATTT	7782
QY	7579	GCTCAGATACAGGCTCAGAAAGTGGTGTGCCCCAGCAAAATCAAACTCCAGTTACCTATC	7638	8485	TATATTGGCTGTGATCGGTGTGAGAAATTTGGTACCATGGCGCTGGCTTGGCATCTTGCAA	8544
Db	6958	-----	6957	7783	TATATTGGCTGTGATCGGTGTGAGAAATTTGGTACCATGGCGCTGGCTTGGCATCTTGCAA	7842
QY	7639	CAAAATTCAGAAAGCAGTCTGTGAGACTCACCAGATTACAGATTGATGTTGTTACAGTGAG	7698	8545	AGTGAGGAGAGCTCATTGATGATGATGCTGTCTCCACAGTGCACAGTGCACAGAGGATGCC	8604
Db	6958	-----	6957	7843	AGTGAGGAGAGCTCATTGATGATGATGCTGTCTCCACAGTGCACAGTGCACAGAGGATGCC	7902
QY	7699	GCAGCCAGTGTGCAAGAGCAGTTGCAAAAGGTTTCAGCAACTCAGGATCAGCAGCAAAAG	7758	8605	ATGACAGTCTCACGCCACTTAACAGAGAGGATTTAGGGGTTGAAGAGGTTGCTCCCT	8664
Db	6958	-----	6957	7903	ATGACAGTCTCACGCCACTTAACAGAGAGGATTTAGGGGTTGAAGAGGTTGCTCCCT	7962
				8665	TCCTTACAGGCCCAATAAGATGGCCTTTTCCTTGAACACAGTAGACCTTAATGATGCA	8724

Db 7963 TCCTTACAGCCCATAGATGGCTTCCCTTGACACAGTAGACCCCTAATGATGA 8022
Qy 8725 CCAGATTATATGTTGTTATTAGGAACCTATGGACCTTGCACCATGGAAGAAGATGA 8784
Db 8023 CCAGATTATATGTTGTTATTAGGAACCTATGGACCTTGCACCATGGAAGAAGATGA 8082
Qy 8785 CAAGAGCATATATGAAGAAGCTGACGGAATTTGTGGCAGATATGACCAAAATTTTGTAT 8844
Db 8083 CAAGAGCATATATGAAGAAGCTGACGGAATTTGTGGCAGATATGACCAAAATTTTGTAT 8142
Qy 8845 AACTGTCCTTACTACATCAACCTGACTCCCATTTTACAGTGTGCAGAAGTTCCTCGAA 8904
Db 8143 AACTGTCCTTACTACATCAACCTGACTCCCATTTTACAGTGTGCAGAAGTTCCTCGAA 8202
Qy 8905 TCATCTTTTGTACAGAAATGAAGGCTTCAAGCTAGCAGGTCTCATACACAACCTG 8964
Db 8203 TCATCTTTTGTACAGAAATGAAGGCTTCAAGCTAGCAGGTCTCATACACAACCTG 8262
Qy 8965 CAGTCTACAGTCTTTAA 8982
Db 8263 CAGTCTACAGTCTTTAA 8280

RESULT 4

ABK84196
ID ABR84196 standard; cDNA; 2673 BP.

XX AC ABR84196;

XX DT 14-AUG-2002 (first entry)

XX DE Human cDNA differentially expressed in granulocytic cells #767.

XX KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.

XX OS Homo sapiens.

XX PN WO200228999-A2.

XX PD 11-APR-2002.

XX PF 03-OCT-2001; 2001WO-US30821.

XX PR 03-OCT-2000; 2000US-237189P.

XX PA (GENE-) GENE LOGIC INC.

XX PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

XX DR WPI; 2002-435328/46.

XX PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -

XX PS Claim 1; SEQ ID No 767; 114pp; English.

XX CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)

CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX XX Sequence 2673 BP; 887 A; 530 C; 678 G; 578 T; 0 other;

Query Match 26.0%; Score 2565.6; DB 24; Length 2673;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 2594; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

Qy 216 GTACGATGACACGAGCGAGCGGTGGAGGAAGAGGAGGACATGCTCCGAGGAGA 275

Db 31 GTACGGGGGGGGGGGGGGGGGGGGAGGAGGAAGAGGAGGACATGCTCCGAGGAGA 90

Qy 276 GGAGGAGGAGGAGCGGCGAGCGCCGAGGAGACCCAGGATTCGAGGACGAGGAGATGA 335

Db 91 GGAGGAGGAGGAGCGGCGAGCGCCGAGGAGACCCAGGATTCGAGGACGAGGAGATGA 150

Qy 336 GATGGAAGAGGACGACGATGACTCCCATTCGCGAGGAGATGGAAGACGACGACGACA 395

Db 151 GATGGAAGAGGACGACGATGACTCCCATTCGCGAGGAGATGGAAGACGACGACGACA 210

Qy 396 CGCCAGTTACTGCACGGAAGACAGCTTCAGGAGCCATAGTACCTACAGCAGCACTCCAG 455

Db 211 CGCCAGTTACTGCACGGAAGACAGCTTCAGGAGCCATAGTACCTACAGCAGCACTCCAG 270

Qy 456 TAGGCGAAACCAAGAGTACATCGGCTCTCTCTCTATATTTGGAGAAACACATCCC 515

Db 271 TAGGCGAAACCAAGAGTACATCGGCTCTCTCTCTATATTTGGAGAAACACATCCC 330

Qy 516 GCCCTTTGAATTTCCCAAGTCCTGAGGATTTAAATGGTGCCTAATGAGCATATAATGA 575

Db 331 GCCCTTTGAATTTCCCAAGTCCTGAGGATTTAAATGGTGCCTAATGAGCATATAATGA 390

Qy 576 TGTCAATGCCATTTTACGAGGTACTTGGCAACTTTGGCACTGTTTGGATTTACTCCTTT 635

Db 391 TGTCAATGCCATTTTACGAGGTACTTGGCAACTTTGGCACTGTTTGGATTTACTCCTTT 450

Qy 636 TCGCTTTGAGACTTTTGTGCAGCTCTGTTGAGCCAGAGCAGTGCACACTCATGGCAGA 695

Db 451 TCGCTTTGAGACTTTTGTGCAGCTCTGTTGAGCCAGAGCAGTGCACACTCATGGCAGA 510

Qy 696 GATGCATGTTGTGCTTTTGAAGCAGTTCGCGTGAAGAAGACACTTCCCAATACTACCTT 755

Db 511 GATGCATGTTGTGCTTTTGAAGCAGTTCGCGTGAAGAAGACACTTCCCAATACTACCTT 570

Qy 756 TGGACCTGCTGATGTGAAGATAGCGTTAATTCACACTGTATTCATAGATGGATGAC 815

Db	571	TGACCTGCTGATCTGAAGATAGCGTTAAATTCACACTGTATTTTCATAGATGGGATGAC	630
QY	816	GTGGCCAGAGTGCTGCGGGTCTACTGTGAGAGTGATAAGGAGTAGTACCATCAGTCTCTTCC	875
Db	631	GTGGCCAGAGTGCTGCGGGTCTACTGTGAGAGTGATAAGGAGTAGTACCATCAGTCTCTTCC	690
QY	876	TTACCAAGAGGCGAGGAGTACCCATATGAGCAGTAGAGAACAGATCAAAAGTTCTACA	935
Db	691	TTACCAAGAGGCGAGGAGTACCCATATGAGCAGTAGAGAACAGATCAAAAGTTCTACA	750
QY	936	GTTCCTAGTCGATCAGTTCTTTACACAAATATTTGCTCCAGAGGAATTCATCTCGAAGG	995
Db	751	GTTCCTAGTCGATCAGTTCTTTACACAAATATTTGCTCCAGAGGAATTCATCTCGAAGG	810
QY	996	GGTGATACAGTATGATGACCATTTGTAGGTTTGTGCACAAACTTTGGGGATTTTCTTTGCTG	1055
Db	811	GGTGATACAGTATGATGACCATTTGTAGGTTTGTGCACAAACTTTGGGGATTTTCTTTGCTG	870
QY	1056	TCAGACATGTTTCAGCAGTATACCATTTTGGAAATGTGTGAAGCCACCTCTTTGAGGAGTGCC	1115
Db	871	TCAGACATGTTTCAGCAGTATACCATTTTGGAAATGTGTGAAGCCACCTCTTTGAGGAGTGCC	930
QY	1116	AGAGCAGGAGTGGCAGTGTGAAGTCTGTGTACACACACAGGTGCCTGGTGTGACTGACTG	1175
Db	931	AGAGCAGGAGTGGCAGTGTGAAGTCTGTGTACACACACAGGTGCCTGGTGTGACTGACTG	990
QY	1176	TCGTCGTGAATCCAAAAAATAAACCATATATTCGACATGAACCTATTGGGATATGATAG	1235
Db	991	TCGTCGTGAATCCAAAAAATAAACCATATATTCGACATGAACCTATTGGGATATGATAG	1050
QY	1236	AAGTCGGAGGAATAACTGGTCTTGAACCGAAGACTCATATAGAGAGATACAGAAAA	1295
Db	1051	AAGTCGGAGGAATAACTGGTCTTGAACCGAAGACTCATATAGAGAGATACAGAAAA	1110
QY	1296	TCAGAAATGAAGAAAAATTTGGTATTTACAGCACAAGGTCACAACTTGCGAGAATTAATGA	1355
Db	1111	TCAGAAATGAAGAAAAATTTGGTATTTACAGCACAAGGTCACAACTTGCGAGAATTAATGA	1170
QY	1356	CTGCTCTAGACAAAGATTAATTTGGGAAGCAGAACTCTGCAAAATTTCTAGAGAAATCGGTGA	1415
Db	1171	CTGCTCTAGACAAAGATTAATTTGGGAAGCAGAACTCTGCAAAATTTCTAGAGAAATCGGTGA	1230
QY	1416	AGAAATCCACCACACATGGACATTAACCTGAAGACCTTGACCAATTAAGGCTCGGGGAGTAA	1475
Db	1231	AGAAATCCACCACACATGGACATTAACCTGAAGACCTTGACCAATTAAGGCTCGGGGAGTAA	1290
QY	1476	CAAACTCCTTTCTGGCGGAGCTTAATGAAGAAATTTTGGAAATCCATAAGAGCCAAAAGGG	1535
Db	1291	CAAACTCCTTTCTGGCGGAGCTTAATGAAGAAATTTTGGAAATCCATAAGAGCCAAAAGGG	1350
QY	1536	AGACATTTGATAATTTTAAAGCCAGAGAAACAGAAAAAGACAAGAAATGAGACTGAGAA	1595
Db	1351	AGACATTTGATAATTTTAAAGCCAGAGAAACAGAAAAAGACAAGAAATGAGACTGAGNA	1410
QY	1596	TGACTCTAAAGATGCTGAGAAAAACAGAGAGAAATTTGAAGACCACTGCTTGAAGAAAG	1655
Db	1411	TGACTCTAAAGATGCTGAGAAAAACAGAGAGAAATTTGAAGACCACTGCTTGAAGAAAG	1470
QY	1656	CAGTCAGCACAACACACAGATGATCACCCTGAGCAAGGAAATCTGAGGAGCCCAACAGA	1715
Db	1471	CAGTCAGCACAACACACAGATGATGACCCTGAGCAAGGAAATCTGAGGAGCCCAACAGA	1530
QY	1716	AGTTGGGGATAAGGTAAGTCTGTGTGACCAATTTTGGCGACAACACAAACAAATGCAAC	1775
Db	1531	AGTTGGGGATAAGGTAAGTCTGTGTGACCAATTTTGGCGACAACACAAACAAATGCAAC	1590
QY	1776	TTTCAGAGAGACTAGTCCCTCTGAAGGGAGGAGCCCTGTGGGGTGTCTCTCAGAAACCC	1835
Db	1591	TTTCAGAGAGACTAGTCCCTCTGAAGGGAGGAGCCCTGTGGGGTGTCTCTCAGAAACCC	1650
QY	1836	CGATAGCAGCACATGGCAGAGAAAGGTCGCATCTGAGCTCCCCCAGGATGTGCCAGA	1895

RESULT 5
AAH15805
ID AAH15805 standard; cDNA; 2057 bp.
XX
AC AAH15805;
XX

Db	1651	CGATAGCAGCAACATGGCGCAGAGAGAAAGGTGGCATCTTGAGCTCCCCCAGGATGTGCCAGA	1710
QY	1896	AGAACTTAACAGACATGTGAGAGCAGTAACACTAGTGTACCACTTACCTTCCATCCAGCC	1955
Db	1711	AGAACTTAACAGACATGTGAGAGCAGTAACACTAGTGTACCACTTACCTTCCATCCAGCC	1770
QY	1956	TAATCTGGAACACAGTAACAGCAGCAGTGAATAAATTTTCCAGAGTGAATCTGCTAA	2015
Db	1771	TAATCTGGAACACAGTAACAGCAGCAGTGAATAAATTTTCCAGAGTGAATCTGCTAA	1830
QY	2016	GCACGTGATGATCCTGAAATGGAGAAAGAAATCTCATACACCTGTCTCTATTTCAGGA	2075
Db	1831	GCACGTGATGATCCTGAAATGGAGAAAGAAATCTCATACACCTGTCTCTATTTCAGGA	1890
QY	2076	AGAGATAGTAGTGATTTTTCACATCGGAGAAAGTCCACCGGGAGCTAAAGTGAATCTCCTGG	2135
Db	1891	AGAGATAGTAGTGATTTTTCACATCGGAGAAAGTCCACCGGGAGCTAAAGTGAATCTCCTGG	1950
QY	2136	AGCTGGAAGAGAGCATCTGGCTCAACTCGAATCATCATCATCATCATCATCATCATCATCAT	2195
Db	1951	AGCTGGAAGAGAGCATCTGGCTCAACTCGAATCATCATCATCATCATCATCATCATCATCAT	2010
QY	2196	CAAACTTAGTCAAGCTCAAGAGCCAGCAGGTGCAGCGCTGCACATGAAGCAAAATAAT	2255
Db	2011	CAAACTTAGTCAAGCTCAAGAGCCAGCAGGTGCAGCGCTGCACATGAAGCAAAATAAT	2070
QY	2256	ATTTAAGGAGGCAAGAGGTACTGGTAGTTAACTCTCAAGGAGAAATTTACGGTTGAG	2315
Db	2071	ATTTAAGGAGGCAAGAGGTACTGGTAGTTAACTCTCAAGGAGAAATTTACGGTTGAG	2130
QY	2316	CACCAAAAAGGAGTATCATGAAAGAAATATCAACAATTTATTTAAATTTGGGTCAAGA	2375
Db	2131	CACCAAAAAGGAGTATCATGAAAGAAATATCAACAATTTATTTAAATTTGGGTCAAGA	2190
QY	2376	AGGGAAGTATCCGCTTACCAACATCAATCTCCACCAATTCATTTGCTTTGAATAAGCA	2435
Db	2191	AGGGAAGTATCCGCTTACCAACATCAATCTCCACCAATTCATTTGCTTTGAATAAGCA	2250
QY	2436	CCAGCACAGAGAGACCATGATAAGAGAAAGGATCTTGACACATAAGTCTGTCTGACTCC	2495
Db	2251	CCAGCACAGAGAGACCATGATAAGAGAAAGGATCTTGACACATAAGTCTGTCTGACTCC	2310
QY	2496	ACGAGAGAGTTCAATGGAACGGTCTGTCTCATGGGTCCAAAGTTCTTACCATATCTAC	2555
Db	2311	ACGAGAGAGTTCAATGGAACGGTCTGTCTCATGGGTCCAAAGTTCTTACCATATCTAC	2370
QY	2556	TCGTGAGACTGACTATCACCAATTAGAAAAACAACATCCCTTCATCCCTTTTTCATCCCAA	2615
Db	2371	TCGTGAGACTGACTATCACCAATTAGAAAAACAACATCCCTTCATCCCTTTTTCATCCCAA	2429
QY	2616	CTGGGCATCACATAGGGCAAAATTTGGATCAAGGCAGTTTCAGATGTGTAGCAAAACCCAG	2675
Db	2430	CTGGGCATCACATAGGGCAAAATTTGGATCAAGGCAGTTTCAGATGTGTAGCAAAACCCAG	2489
QY	2676	ATTTCGATTTGGCTTTAGCCATTTTGGAGTGTGAGTTAAACAGTTTGTGATGCTTACCAAT	2735
Db	2490	ATTTCGATTTGGCTTTAGCCATTTTGGAGTGTGAGTTAAACAGTTTGTGATGCTTACCAAT	2549
QY	2736	ATGGCGAGAAATTTTAGGACATACCAGGTTTACACCGGATGACATCAATTTGAAGAGAGA	2795
Db	2550	ATGGCGAGAAATTTTAGGACATACCAGGTTTACACCGGATGACATCAATTTGAAGAGAGA	2609
QY	2796	AAAGGAGAAAGTCAAAAAAAGAGAGAAAGGAGAAAGGAGAGAG 2839	
Db	2610	AAAGGAGAAAGTCAAAAAAAGAGAGAAAGGAGAGAGAGG 2653	

Db 236 GTCAGAAATGGTACCATGGCGCTGGCTTGGCATCTTGCAGAGTGAGGAGAGCTCAATG 295
 QY 8564 ATGAGTATGCTCTGCACAGTGCAGTCAACAGAGATGCCATGACAGTGCCTCACGCCAC 8623
 Db 296 ATGAGTATGCTCTGCACAGTGCAGTCAACAGAGATGCCATGACAGTGCCTCACGCCAC 355
 QY 8624 TAACAGAGAAGGATTATGAGGGGTTGAAGAGGGTGCCTCGTTCCTTACAGGCCCAATAAGA 8683
 Db 356 TAACAGAGAAGGATTATGAGGGGTTGAAGAGGGTGCCTCGTTCCTTACAGGCCCAATAAGA 415
 QY 8684 TGGCCTGGCCCTTCCCTTGAACACAGTAGACCCCTAATGATGCACAGATTAATATGGTGTTA 8743
 Db 416 TGGCCTGGCCCTTCCCTTGAACACAGTAGACCCCTAATGATGCACAGATTAATATGGTGTTA 475
 QY 8744 TTAAGGAACCTATGGACCTTGCACCATGGAAGAAAGAGTACAAAGACGATATTAAGAA 8803
 Db 476 TTAAGGAACCTATGGACCTTGCACCATGGAAGAAAGAGTACAAAGACGATATTAAGAA 535
 QY 8804 AGCTGACGGAATTTGTGGCAGATATGACCAAAATTTTGTGATACTGTCGTTACTCAATC 8863
 Db 536 AGCTGACGGAATTTGTGGCAGATATGACCAAAATTTTGTGATACTGTCGTTACTCAATC 595
 QY 8864 CAAGTACTCCCATTTTACAGTGCAGAAAGTCTCGAATCATCTTTGTACAGAAAT 8923
 Db 596 CAAGTACTCCCATTTTACAGTGCAGAAAGTCTCGAATCATCTTTGTACAGAAAT 655
 QY 8924 TGAAGGCTTCAAAGCTAGCAGTCTCATACCAAACTGCAGTCTACAGTCTCTTAA 8983
 Db 656 TGAAGGCTTCAAAGCTAGCAGTCTCATACCAAACTGCAGTCTACAGTCTCTTAA 715
 QY 8984 GTTCAGGCTGTAACTTAACATTAACACAGCAAGAACTGGTGTGCTGAACATTTTAA 9043
 Db 716 GTTCAGGCTGTAACTTAACATTAACACAGCAAGAACTGGTGTGCTGAACATTTTAA 775
 QY 9044 ATTAAGAGCCAGATGTTTTAGTCAGGCTATCCTGACAGACTTGACCTAAACTTCGTT 9103
 Db 776 ATTAAGAGCCAGATGTTTTAGTCAGGCTATCCTGACAGACTTGACCTAAACTTCGTT 835
 QY 9104 TTTATTGGTCAACAGTCCCAATATATTTCTTGCCCAATTTTGTCCCAACGCAAGAAA 9163
 Db 836 TTTATTGGTCAACAGTCCCAATATATTTCTTGCCCAATTTTGTCCCAACGCAAGAAA 895
 QY 9164 AAGCAAGTCAACAGACCAATATCTTGTCAAGATCAGATGGTTTTACTA-TTGTGCGAG 9222
 Db 896 AAGCAAGTCAACAGACCAATATCTTGTCAAGATCAGATGGTTTTACTA-TTGTGCGAG 955
 QY 9223 AAGCAGAAACTTTGTTTATT--GAAAAAAGAAAAAGAAAGCAAGAAAAAGATA 9280
 Db 956 AAGCAGAAACTTTGTTTATTAAAGAAAAAAGAAAGCAAGAAAAAGCAATGATA 1015
 QY 9281 CTATGGGTCAGTGTAACTCCATGGAAAGCCAGCTGCTGCTTCAAGTGAAGAGCTGG 9340
 Db 1016 CTGTGGGTCAGTGTAACTCCATGGAAAGCCAGCTGCTGCTTCAAGTGAAGAGCTGG 1075
 QY 9341 TTTAGAGTCTCAGAAAACTTTGACTGATTTATTTATTTGTTGCAAAAAGAGCGTTT 9400
 Db 1076 TTTAGAGTCTCAGAAAACTTTGACTGATTTATTTATTTGTTGCAAAAAGAGCGTTT 1135
 QY 9401 TTTATTGCTGCCCTCATTTGTCAGTAACTAGTATTTTCTTATAAAATCCAGCCCGGTTA 9460
 Db 1136 TTTATTGCTGCCCTCATTTGTCAGTAACTAGTATTTTCTTATAAAATCCAGCCCGGTTA 1195
 QY 9461 CATATAAT-CATCTGTATCTTATC-ATGATTCCTGTAGGTAAGTACACAGACCTCT 9518
 Db 1196 CATATAATCCATCCATATCTTATCAATGATTCCTGTAGTAAAAGTACAGACACCTCT 1255
 QY 9519 AGATGCTTTCTTCTATGAAGAGCTGCTATGTACACATGTGCACACACACAACT 9578
 Db 1256 AGATGCTTTCTTCTATGAAGAGCTGCTATGTACACATGTGCACACACACAACT 1315
 QY 9579 GGAATCAACATGAGTTTATTTGTTATGGTAGATTAATAAATTAAGCTTGCATAAAGGTT 9638
 Db 1316 GGAATCAACATGAGTTTATTTGTTATGGTAGATTAATAAATTAAGCTTGCATAAAGGTTG 1375

QY 9639 GCTAAGTGGTCTCTGGGCTACAGACTCTGTTGGCTTGAATATACAGTACAAATTTGCA 9698
 Db 1376 GCTAAGTGGTCC-TGGACTACAGACTCTGGTGAATATAACAGTACAAATTTGCA 1434
 QY 9699 ATTACTCTCCACAGGCTAAAGTGAGTAAATCTATTTGAAGGTATCTTTGTTGTAACA 9758
 Db 1435 ATTACTCCACAGGCTTGAATGAGTAAATCTATTTGAAGGTATCTTTGTTGTAACA 1494
 QY 9759 TTTGTGAGATCTAAATTTTCTTTTGTATTAATAATTCAACTATGG 9805
 Db 1495 TTTGTGAGATCTAAATTTTCT-TTTGTATTAATAATTCAACTATGG 1540

RESULT 7

AAX30155
 ID AAX30155 standard; DNA; 1032 BP.
 XX
 AC AAX30155;
 XX
 DT 18-JUN-1999 (first entry)
 XX
 DE Human secreted protein gene 11.
 XX
 KW Human; secreted protein; cancer; tumour; developmental abnormality;
 KW foetal deficiency; blood disorder; immune system disorder; inflammation;
 KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
 KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
 KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
 KW digestive disorder; endocrine disorder; infection; AIDS; ss.
 XX
 OS Homo sapiens.
 OS
 PN WO9910363-A1.
 XX
 PD 04-MAR-1999.
 XX
 PF 27-AUG-1998; 98WO-US17709.
 XX
 PR 29-AUG-1997; 97US-0056271.
 PR 29-AUG-1997; 97US-0056073.
 PR 29-AUG-1997; 97US-0056247.
 PR 29-AUG-1997; 97US-0056270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fan P, Kyaw H, Rosen CA, Ruben SM, Wei YF;
 XX
 DR WPI: 1999-190585/16.
 DR P-PSDB; AAY04303.
 XX
 PT New isolated human genes and the secreted polypeptides they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 1; Page 138; 170pp; English.
 XX
 CC AAX30145 to AAX30173 represent 29 isolated human secreted protein genes.
 CC AAY04293 to AAY04321 represent the secreted proteins encoded by the 29
 CC human genes. The genes and their corresponding secreted polypeptides are
 CC useful for preventing, treating or ameliorating medical conditions, be
 CC e.g. by protein or gene therapy. Also pathological conditions can be
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new genes. Specific
 CC uses are described for each of the 29 genes, based on which tissues they
 CC are most highly expressed in, and include developing products for the
 CC diagnosis or treatment of cancer, tumours, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
 CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
 CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
 CC disorders, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.

Qy	5862	TATTATTGAAACCTGGGTAGCAGAAAGAAAGAACTGGAAATTGTCGAGATCAGGCATTGCG	5921
Db	182	TATTATTGAAACCTGGGTAGCAGAAAGAAAGAACTGGAAATTGTCGAGATCAGGCATTGCG	241
Qy	5922	TGAGAGAGTGGAGAAAGAAAAAGGCACAAAGCAGTTGAGCAACAGGCTAAGAAACGACTGGA	5981
Db	242	TGAGAGAGTGGAGAAAGAAAAAGGCACAAAGCAGTTGAGCAACAGGCTAAGAAACGACTGGA	301
Qy	5982	GCAGCAAGCCGACAGTGAATGCCAATCTCCACTACTTCCCAACAAGCAGTACAACGAG	6041
Db	302	GCAGCAAGCCGACAGTGAATGCCAATCTCCACTACTTCCCAACAAGCAGTACAACGAG	361
Qy	6042	CACCATCTCTCCAGCACAGAAAGTTATGTGTGGCCCCCATTAAGTGGCTCAGTTACAACCTGG	6101
Db	362	CACCATCTCTCCAGCACAGAAAGTTATGTGTGGCCCCCATTAAGTGGCTCAGTTACAACCTGG	421
Qy	6102	AACCAAAATGGTACTACTTAAGTTGGATCTCCAGCTACAGTACAGTTCCAACAAAA	6161
Db	422	AACCAAAATGGTACTACTTAAGTTGGATCTCCAGCTACAGTACAGTTCCAACAAAA	481
Qy	6162	CAAGAACTTTTCATCAAAACCTTTGCTACATGGGTTAAGCAAGGCCAGTCAAAATTCAGGCGT	6221
Db	482	CAAGAGCTTTTCATCAAAACCTTTGCTACATGGGTTAAGCAAGGCCAGTCAAAATTCAGGCGT	541
Qy	6222	TGTTCAAGTACAGCAGAAAGTCCTGGGTATCATTTCCATCAAGTACAGGTACAGTACAGCA	6281
Db	542	TGTTCAAGTACAGCAGAAAGTCCTGGGTATCATTTCCATCAAGTACAGGTACAGTACAGCA	601
Qy	6282	AACCTTTACTTCATTCCAGCCCCAGGACAGCAACAGTCACAAATTAGGCCCAATACCTCAGG	6341
Db	602	AACCTTTACTTCATTCCAGCCCCAGGACAGCAACAGTCACAAATTAGGCCCAATACCTCAGG	660
Qy	6342	CTCTGGAGGAACCAAGCAATTCACAAGTAATCACAGGCGCTCAGATTCGCCCTGGTAT	6401
Db	661	CTCTGGAGGAACCAAGCAATTCACAAGTAATCACAGGCGCTCAGATTCGCCCTGGTAT	720
Qy	6402	GACCGTCATTAGAACACCACTCCAACAGTCAACACTAGGAAGGCAA-TTATTTCGAACAC	6460
Db	721	GACCGTCATTAGAACACCACTCCAACAGTCAACACTAGGAAGGCAATTNTTCGAACAC	780
Qy	6461	CTGTGATGGTACGCCAGGTGCTCCTCAGCAAGTG	6495
Db	781	CTGTGATGGTACGCCAGGTGCTCCTCAGCAAGG	815

RESULT 9	
ABK47549	
ID	ABK47549 standard; DNA; 735 BP.
XX	
XX	
AC	ABK47549;
XX	
XX	
DT	18-JUN-2002 (first entry)
XX	
DE	DNA encoding SCAN/KRAB protein SKAT-2 binding protein, 7A1.
XX	
KW	SCAN/KRAB protein associated with a Th2 phenotype; SKAT-2; asthma;
KW	allergy; hay fever; atopic dermatitis; allergic rhinitis;
KW	Alzheimer's disease; neurological disorder; apoptosis;
KW	T-helper response; vaccination; immune response; autoimmune disease;
KW	infection; cancer; kruppel-like zinc finger protein; 7A1; gene; ds;
KW	human.

XX	Homo sapiens.
OS	
XX	
XX	
PN	GB2364051-A.
XX	
XX	
XX	16-JAN-2002.
PD	
XX	
XX	
PF	06-APR-2000; 2000GB-0008549.
XX	
XX	
PR	06-APR-2000; 2000GB-0008549.
XX	

(GLAXO) GLAXO GROUP LTD.

Blanchard AD, Champion B, Page KR;
WPI; 2002-306794/35.
P-PSDB; AAU78845.

New SKAT-2 polypeptide transcription factor, useful for treating or preventing e.g., allergy and Alzheimer's disease, also related nucleic acid, antibodies and modulators -

Disclosure; Page 37-38; 47pp; English.

The invention describes an isolated SCAN/KRAB protein associated with a Th2 phenotype (SKAT-2) polypeptide (I). (I), especially where expressed in recombinant cells, is used to identify agents that modulate SKAT-2 activity, and may be useful therapeutically. The agents, also (I) and the polynucleotides (II) that encode it, are useful for treatment and prevention of diseases responsive to SKAT-2 modulation, specifically asthma; allergies (hay fever, atopic dermatitis; allergic rhinitis etc.); Alzheimer's disease (and other neurological disorders, particularly where related to apoptosis); conditions involving T-helper responses, and apoptosis; also in vaccination against an antigen to enhance/repress Th2 or humoral responses, to reduce development of a Th1 phenotype, and to manipulate the immune response in autoimmune diseases, infections and cancer. (II) is also used for recombinant production of (I): as source of antisense therapeutics and as primers and probes, e.g. for diagnostic detection of mutations and for monitoring SKAT-2 expression in association with disease. This sequence encodes the peptide 7AI, that binds specifically to the kruppel-like zinc finger protein, SKAT-2.

Sequence 735 BP: 242 A; 143 C; 153 G; 197 T; 0 other;

Query Match	7.58;	Score 735;	DB 24;	Length 735;
Best Local Similarity	100.0%;	Prod. No. 4.8e-136;		
Matches 735;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	8533	GGCATCTTGGCAAAGCTGAGGAGAGCTCATTTGATGAGTATGCTCTGCACAGTGCCAGTGCA	8592	
Db				
Qy	1	GGCATCTTGGCAAAGCTGAGGAGAGCTCATTTGATGAGTATGCTCTGCACAGTGCCAGTGCA	60	
Db				
Qy	8593	ACAGAGGATGCCATGACAGTGCCTCACGCCACTAACACAGAGAAGGATTATGAGGGTTGAAG	8652	
Db				
Qy	61	ACAGAGGATGCCATGACAGTGCCTCACGCCACTAACACAGAGAAGGATTATGAGGGTTGAAG	120	
Db				
Qy	8653	AGGTTGCTCGTTCCTTACAGGCCCATTAAGATGGCCTTGGCCTTTCCTTGAACCAAGTAGAC	8712	
Db				
Qy	121	AGGTTGCTCGTTCCTTACAGGCCCATTAAGATGGCCTTGGCCTTTCCTTGAACCAAGTAGAC	180	
Db				
Qy	8713	CCTAATGATGCACCAGAGATTATTATGGTGTATTAAAGGAACCTATGGGACCTTGGCCACCATTG	8772	
Db				
Qy	181	CCTAATGATGCACCAGAGATTATTATGGTGTATTAAAGGAACCTATGGGACCTTGGCCACCATTG	240	
Db				
Qy	8773	GAAGAAAGAGTACAAAGACGATATTATCAAAAGCTGACGGAAATTTGGCAGATATGACC	8832	
Db				
Qy	241	GAAGAAAGAGTACAAAGACGATATTATCAAAAGCTGACGGAAATTTGGCAGATATGACC	300	
Db				
Qy	8833	AAAATTTTGTAACTGCTGTTACTACAATCCAAAGTGACTCCCATTTTACCAGTGTGCA	8892	
Db				
Qy	301	AAAATTTTGTAACTGCTGTTACTACAATCCAAAGTGACTCCCATTTTACCAGTGTGCA	360	
Db				
Qy	8893	GAAGTTCTCGAATCATTTTGTACAGAAATTTGAAGGCTTCAAGCTACGAGTCTCAT	8952	
Db				
Qy	361	GAAGTTCTCGAATCATTTTGTACAGAAATTTGAAGGCTTCAAGCTACGAGTCTCAT	420	
Db				
Qy	8953	AACAACAAACTGCAGTCTACAGCTTCTTAAAGTTTCAGCGTGTAACTCAACATAAAACAC	9012	
Db				
Qy	421	AACAACAAACTGCAGTCTACAGCTTCTTAAAGTTTCAGCGTGTAACTCAACATAAAACAC	480	
Db				
Qy	9013	AGCAAGAACTGTGGTTGTCTGAACFATTTTTAAATTAAGAGCCAGATGTTTTAGTCAGGC	9072	
Db				
Qy	481	AGCAAGAACTGTGGTTGTCTGAACFATTTTTAAATTAAGAGCCAGATGTTTTAGTCAGGC	540	
Db				

QY	9073	TATCCTGACAGACTTCGACCTAAACCTTCGTTTTTATTTGGTCATTAACAGTCCAATTATATT	9133
Db	541	TATCCTGACAGAACTTGACCTAAACCTTCGTTTTTATTTGGTCATTAACAGTCCAATTATATT	600
QY	9133	CTTGGCCAAATTTGTCTCCACGGCAAGAAAAAGCAAAAGTCAACGACACCATTTATCTTGT	9192
Db	601	CTTGGCCAAATTTGTCTCCACGGCAAGAAAAAGCAAAAGTCAACGACACCATTTATCTTGT	660
QY	9193	CAAGATCAGATGGTTTTTACTATTGTGGCAGACGAGAAAAACTTTGTTATTGAAAAAAA	9252
Db	661	CAAGATCAGATGGTTTTTACTATTGTGGCAGACGAGAAAAACTTTGTTATTGAAAAAAA	720
QY	9253	AAGAAAAAGAAAGCA 9267	
Db	721	AAGAAAAAGAAAGCA 735	
RESULT 10			
AAC99129			
ID	AAC99129 standard; cDNA; 1408 BP.		
XX	AAC99129;		
XX	09-MAR-2001 (first entry)		
XX	Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:357.		
DE	Human; pancreas; pancreatic cancer; pancreatic cancer antigen;		
KW	detection; diagnosis; identification; cytostatic; neuroprotective;		
KW	neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;		
KW	antiinflammatory; cardiant; gene therapy; chromosome mapping;		
KW	linkage analysts; tissue identification; tissue typing; forensic;		
KW	neural; immune system; muscular; reproductive; gastrointestinal;		
KW	pulmonary; cardiovascular; renal; proliferative; ss.		
OS	Homo sapiens.		
XX	WO200055320-A1.		
XX	21-SEP-2000.		
XX	08-MAR-2000; 2000WO-US05989.		
XX	12-MAR-1999; 99US-0124270.		
XX	(HUMA-) HUMAN GENOME SCI INC.		
XX	Rosen CA, Ruben SM;		
PI	WPI; 2000-579444/54.		
XX	P-PSDB; AAB54364.		
XX	New nucleic acid that is a pancreatic cancer antigen for preventing,		
PT	treating, or ameliorating a medical condition, particular pancreatic		
PT	cancer, or for use in assays for diagnosing a pathological condition -		
XX	Claim 1; Page 779-780; 1379pp; English.		
XX	AAC98773 to AAC99231 encode the human pancreatic cancer associated		
CC	proteins, called pancreatic cancer antigens, given in AAB54008 to		
CC	AAB54466. The human pancreatic cancer antigens have cytostatic,		
CC	neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,		
CC	gynaecological, cardiant and antiinflammatory activities, and can be used		
CC	in gene therapy. The polynucleotide and proteins can be used for		
CC	preventing, treating, or ameliorating a medical condition or in assays		
CC	for diagnosing a pathological condition or a susceptibility to one in a		
CC	subject. Binding partners to the proteins and the activity of the		
CC	proteins can be identified. The pancreatic cancer antigens can be used to		
CC	detect, treat or prevent pancreatic disorders, especially cancer.		
CC	Agonists and antagonists to the antigens can be screened for. The		
CC	pancreatic cancer antigen polynucleotides can be used to design nucleic		
CC	acid hybridisation probes that can be used in chromosome mapping, linkage		
CC	analysis, tissue identification and/or typing and a variety of forensic		

CC	and diagnostic methods. The proteins can be used to generate antibodies
CC	which are used to purify, detect and target the polypeptides, including
CC	both in vivo and in vitro diagnostic and therapeutic methods. The
CC	proteins can be used to treat or prevent neural, immune system, muscular,
CC	reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC	proliferative disorders. AAC99232 to AAC99240 and AAC99467 represent
CC	sequences used in the exemplification of the present invention.
XX	
SQ	Sequence 1408 BP; 468 A; 303 C; 314 G; 311 T; 12 other;
	Query Match 6.8%; Score 673.8; DB 21; Length 1408;
	Best Local Similarity 97.1%; Pred. No. 8.7e-124;
	Matches 675; Conservative 9; Mismatches 11; Indels 0; Gaps 0;
Oy	2083 GTAGGTGATTTCACATCGGAGAAGTCCACGGGGAGCTAAGTGAATCTCCTGCAGCTGGA 2142
Dd	276 GTAGGTGATTTCAAATCGGAGAAGTCCAACGGGGAGCTAAGTGAATCTCCTGCAGCTGGA 335
Oy	2143 AAAGGAGCATCTGGCTCAACTCGAATCATCACAGATTCGCGAATCCAGATACCAAACTT 2202
Dd	336 AAAGGAGCATCTGGCTCAACTCGAATCATCACAGATTCGCGAATCCAGATACCAAACTT 395
Oy	2203 AGTCAGCTGAAGACCAGCAGGTGGCAGCGCTGCACATGAACAATAAATTTATTTAAG 2262
Dd	396 AGTCAGCTGAAGACCAGCAGGTGGCAGCGCTGCACATGAACAATAAATTTATTTAAG 455
Oy	2263 GAGGGCAAAGAGTGACTGGTAGTTAACTCTCAAGGAGAAATTTACGGTTTGACACCAAAA 2322
Dd	456 GAGGGCAAAGAGTGACTGGTAGTTAACTCTCAAGGAGAAATTTACGGTTTGACACCAAAA 515
Oy	2323 AAGGAAGTGCATCATGAAGGAAATATCAACAATTTATTTAAATTTGGTCAAGAAAGGAAG 2382
Dd	516 AAGRAAGTGRTCATGAAGGAARATATCAACAATTTATTTAAATTTGGTCAAGAAAGGAAG 575
Oy	2383 TATCGGCTTACACAAATCAATTAATCTACCAATTTCAATTTGCTTTGAATAGCACAGCAC 2442
Dd	576 TATCGGCTTACACAAATCAATTAATCTACCAATTTCAATTTGCTTTGAATAGCACAGCAC 635
Oy	2443 AGAAGAACCATGATAGAGAAGGCATCTTTGCACATAAGTTCTGTCTGACTCCAGCAGGA 2502
Dd	636 AGAAGAACCATGATAGAGAAGGCATCTTTGCACATAAGTTCTGTCTGACTCCAGCAGGA 695
Oy	2503 GAGTTCAATTTGGAACGGTTCTGTCCATGGGTGCCAAAGTTCTTACCATATCTACTCTGAGA 2562
Dd	696 GAGTTCAATTTGGAACGGTTCTGTCCATGGGTGCCAAAGTTCTTACCATATCTACTCTGAGA 755
Oy	2563 CTGACTATACCAATTTAGAAAAACACATCCCCTTTCATCTCTTTTTCATCCCACTGGGCA 2622
Dd	756 CTGACTATACCAATTTAGAAAAACACATCCCCTTTCATCTCTTTTTCATCCCACTGGGCA 815
Oy	2623 TCACATAGGCAAAATTTGGAGTGCAGGAGTTACAGTGTAGCAAAACCCAGAGAAATTTGCA 2682
Dd	816 TCACATAGGCAAAATTTGGAGTGCAGGAGTTACAGTGTAGCAAAACCCAGAGAAATTTGCA 875
Oy	2683 TTGGCTTTAGCCATTTTGGAGTGCAGGTTTAAACCCAGTTGTGTATGTACCAATATGGCGA 2742
Dd	876 TTGGCTTTAGCCATTTTGGAGTGCAGGTTTAAACCCAGTTGTGTATGTACCAATATGGCGA 935
Oy	2743 GAATTTTTAGGACATACCAAGGTTTACACCGGATGAC 2777
Dd	936 GAATTTTTAGGACATACCAAGGTTTACACCGGATGAC 2777
RESULT 11	
AAI94524	ID AAI94524 standard; cdna; 750 BP.
XX	AAI94524;
XX	AAI94524;
XX	13-NOV-2001 (first entry)
XX	Human neuroblastoma expressed polynucleotide SEQ ID NO 599.

KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
 XX Homo sapiens.
 OS WO200166719-A1.
 PN 13-SEP-2001.
 PD 02-MAR-2001; 2001WO-JP01629.
 PE 07-MAR-2000; 2000JP-0159195.
 PR (CHIB-) CHIBA PREFECTURE.
 XX (HISM) HISAMITSU PHARM CO LTD.
 PA Nakagawara A;
 PI WPI; 2001-565584/63.
 DR Nucleic acids originating in gene expressed in human neuroblastoma,
 XX useful as probe or primer in diagnosing prognosis of human
 PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
 PT for anti-cancer agents -
 XX
 PS Claim 1; Page 480; 2979pp; Japanese.
 XX The invention relates to novel genes (AAI93926-AAI97963) expressed in
 CC human neuroblastoma. The nucleic acids are applicable as a probe or
 CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
 CC susceptibility indicators or tumour markers for anti-cancer agents. The
 CC gene information for diagnosing prognosis is related to factors similar
 CC to that for N-myc and TrkA genes.
 XX
 XX Sequence 750 BP; 271 A; 108 C; 169 G; 183 T; 19 other;
 SQ

Query Match 6.4%; Score 627.4; DB 22; Length 750;
 Best Local Similarity 94.9%; Pred. No. 1.1e-114;
 Matches 674; Conservative 0; Mismatches 32; Indels 4; Gaps 3;

QY 4005 TGACAGAGATGCCACCTCTGTCAAGCAATGGACTTTGAAGGAAACCTGGGATGGA 4064
 DB 42 TACAGAGATCGGGTCTCTGTCAAGCAATGGACTTTGAAGGATTTCTGGGATGGA 101
 QY 4065 CTCTGAATCTAATAGCACTTTGGAAAATAGTTCTGTATACCGTGTCTATTCAGTAGCAG 4124
 DB 102 CTCTGAAGNACCCCTCTTTGGAAAATATCTGTATACCGTGTCTATTCAGTAGCAG 161
 QY 4125 TGAAGAAGATATGATTGTTTCAGCAATAGCAATGAAGCAATTTCTGAACAGTTTCAGAACTCG 4184
 DB 162 TGAAGAAGATATGATTGTTTCAGCAATAGCAATGAAGCAATTTCTGAACAGTTTCAGAACTCG 221
 QY 4185 AGAACAGATGTTGAAGTCTTGGAGCCGTTAAAGTGTGAGTTGGTTCTGTGAGTCCAC 4244
 DB 222 AGAACAGATGTTGAAGTCTTGGAGCCGTTAAAGTGTGAGTTGGTTCTGTGAGTCCAC 281
 QY 4245 TGGAACTGTGAGGACAGGCTGCCGGTCAAGGGACTGAAGCAATAGTTGTAACCAACG 4304
 DB 282 TGGAACTGTGAGGACAGGCTGCCGGTCAAGGGACTGAAGCAATAGTTGTAACCAACG 341
 QY 4305 TCAGCAGAAATATAGAGGAGAGACCACTGTAATTAATAGTATGATCAATTAAGCTAAA 4364
 DB 342 TCAGCAGAAATATAGAGGAGAGACCACTGTAATTAATAGTATGATCAATTAAGCTAAA 401
 QY 4365 AAATACCACTGACAAAAAGATAATGAAATCGAGAGTCTGAAAGAAAGGACAGAGAAC 4424
 DB 402 AAATACCACTGACAAAAAGATAATGAAATCGAGAGTCTGAAAGAAAGGACAGAGAAC 461
 QY 4425 AAGTACATTTCAATTAATGAAAGAGATAATAAACCCAAAATATATTTGAAGGTGAATG 4484
 DB 462 AAGTACATTTCAATTAATGAAAGAGATAATAAACCCAAAATATATTTGAAGGTGAATG 521
 QY 4485 CTGAAACAAATTTCTGAGAGTAGTAGTGAAGTGTGATGTTGAACCAAGGTTAATAA 4544

Db 522 CTTGAAAGAAATTTCTGAGAGTAGTA-TAAGTGGTAAATGTTGAACCAAGGTTAATAA 580
 QY 4545 TATAAATAAAATATCCCTGAGAAATGATATTAATAATCATTTGACTGTTAAAGAAATCTGCTAT 4604
 Db 581 TATANATAAAATATCCCTGAGAAATGATATTAATAATCATTTGACTGTTAAAGAAATCTGCTAT 640
 QY 4605 AAGGCCATTCATTAAATGGTGTATGTCATCATGG--AAGATTTTAAATGAAGAAACAGCTCC 4662
 Db 641 AAGGCCATTCATTAAATGGTGTATGTCCTCATGGGAAGATTTTNTANTGAAGAAACAGCTCC 700
 QY 4663 GAAACAAAATCCGATTTGCTGAGTCTTCTCAGATG-CTGAAGGTAACCTACC 4711
 Db 701 GAAACAAAATCCGATTTGCTCAGATGCTCTCANATGCCCTGAAGGTAACCTACC 750

RESULT 12
 ACA57390
 ID ACA57390 standard; cDNA; 574 BP.
 XX ACACA57390;
 XX ACACA57390;
 DT 10-JUN-2003 (first entry)
 XX Human adipocyte Selected Interacting domain, SID, cDNA #477.
 DE Human; ss; gene; prey; adipocyte; SID; selected interacting domain;
 KW anorectic; antidiabetic; protein-protein interaction; diabetes;
 KW yeast 2-hybrid assay; metabolic disorder; obesity.
 OS Homo sapiens.
 XX WO200286122-A2.
 PN 31-OCT-2002.
 PD 14-MAR-2002; 2002WO-EP03768.
 PF 14-MAR-2001; 2001US-275734P.
 PR (HYBR-) HYBRIGENICS.
 PA Legrain P, Daviet L;
 PI WPI; 2003-103412/09.
 DR P-PSDB; ABU70846.
 XX
 XX New complex between two interacting proteins in adipocyte cells, useful
 PT for identifying selected interacting domains that modulate protein
 PT interactions, or for preventing or treating metabolic disorders such as
 PT obesity or diabetes -
 XX
 PS Claim 7; Page 267-268; 382pp; English.
 XX
 XX The invention relates to a complex between two interacting proteins in
 CC adipocyte cells, given in the specification. The proteins are identified
 CC by selecting a bait protein from a known adipocyte marker and then
 CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
 CC members of an adipocyte cDNA library. The proteins are designated SID
 CC (RTM) (selected interacting domains) proteins. Also included are a
 CC polynucleotide encoding a polypeptide in the adipocyte cells, a
 CC recombinant host cell expressing at least one of the interacting
 CC polypeptides of the complex, selecting a modulating compound in adipocyte
 CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
 CC sequences given in the specification (including its fragment or variant),
 CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
 CC given in the specification (including its fragment or variant), a vector
 CC comprising the SID (RTM) polynucleotide, a recombinant host cell
 CC comprising the vector, a protein chip comprising the polypeptides and
 CC a record comprising all or part of the data, listed in the specification.
 CC The complex, polypeptides, polynucleotides and compounds are
 CC useful for preventing or treating metabolic disorders such as obesity
 CC or diabetes. The polynucleotides are useful as probes or primers. The
 CC complex is particularly useful for identifying selected interacting

```
CC domains (SID (RTM)) for screening drugs that modulate the protein
CC interaction, thus exhibiting the therapeutic effect. The present
CC sequence encodes a SID (prey) protein of the invention.
XX
SQ Sequence 574 BP; 179 A; 123 C; 134 G; 138 T; 0 other;

Query Match          5.8%; Score 574; DB 25; Length 574;
Best Local Similarity 100.0%; Pred. No. 46-104;
Matches 574; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8301 ACACACAGGCTTCTCTCCACGCCACCTTACCTGTGCTTCCAGAGAGAGCGGGA 8360
Db 1 ACACACAGGCTTCTCTCCACGCCACCTTACCTGTGCTTCCAGAGAGAGCGGGA 60
QY 8361 AGAGGAAAAGACTCCAGTCAAAAGTCCAGAAAAGAAATGATCTACTACCTCAA 8420
Db 61 AGAGGAAAAGACTCCAGTCAAAAGTCCAGAAAAGAAATGATCTACTACCTCAA 120
QY 8421 GGAACCTAAGAGGACACAAAGCTTTACTGTATCTGTAAACGCCCTTATGATGAATCTAA 8480
Db 121 GGAACCTAAGAGGACACAAAGCTTTACTGTATCTGTAAACGCCCTTATGATGAATCTAA 180
QY 8481 ATTTTATATTGGCTGTGATCGGTGTCAGAAATGGTACCATGGCGCTGCGTTGGCATCTT 8540
Db 181 ATTTTATATTGGCTGTGATCGGTGTCAGAAATGGTACCATGGCGCTGCGTTGGCATCTT 240
QY 8541 GCAAAGTGAGGACAGAGCTCATGATGAGTATGCTGTCTCCACAGTGCCAGTCAACAGAGGA 8600
Db 241 GCAAAGTGAGGACAGAGCTCATGATGAGTATGCTGTCTCCACAGTGCCAGTCAACAGAGGA 300
QY 8601 TGCCATGACAGTGTCTACGCCACTACACAGAGAGTATGAGGGGTTGAAGAGGTGCT 8660
Db 301 TGCCATGACAGTGTCTACGCCACTACACAGAGAGTATGAGGGGTTGAAGAGGTGCT 360
QY 8661 CCGTTCCTTACAGGCCCATAAAGATGGCTTGGCTTTTCCCTTGAACCACTAGACCCCTAATGA 8720
Db 361 CCGTTCCTTACAGGCCCATAAAGATGGCTTGGCTTTTCCCTTGAACCACTAGACCCCTAATGA 420
QY 8721 TGACACAGATTTATATGGTGTATTAAGGAACCTATGGACCTTGGCCACCATGGAAGAAAG 8780
Db 421 TGACACAGATTTATATGGTGTATTAAGGAACCTATGGACCTTGGCCACCATGGAAGAAAG 480
QY 8781 AGTACAAAGACGATATTTATGAAAGCTGACGGAATTTGGCAGATATGACCAAAATTTT 8840
Db 481 AGTACAAAGACGATATTTATGAAAGCTGACGGAATTTGGCAGATATGACCAAAATTTT 540
QY 8841 TGATAACTGTCGTACTACAAATCCAAAGTGAATCC 8874
Db 541 TGATAACTGTCGTACTACAAATCCAAAGTGAATCC 574

RESULT 13
AAH17323
ID AAH17323 standard; cDNA; 2123 BP.
AC AAH17323;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:16734.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
PF 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
PR
```

```
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
PS Claim 8; SEQ ID 16734; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
SQ Sequence 2123 BP; 679 A; 439 C; 485 G; 520 T; 0 other;

Query Match          5.8%; Score 574; DB 22; Length 2123;
Best Local Similarity 99.1%; Pred. No. 6.7e-104;
Matches 577; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8664 TTCCTTACAGGCCCATAAAGATGGCTTGGCTTTTCCCTTGAACCACTAGACCCCTAATGATGC 8723
Db 1542 TTCATTATAGGCCCATAAAGATGGCTTGGCTTTTCCCTTGAACCACTAGACCCCTAATGATGC 1601
QY 8724 ACCAGATTATTTATGGTGTATTAAAGAAACCTATGGACCTTGCACCATGGAAGAAAGAGT 8783
Db 1602 ACCAGATTATTTATGGTGTATTAAAGAAACCTATGGACCTTGCACCATGGAAGAAAGAGT 1661
QY 8784 ACAAGACGATATTTATGAAAGCTGACGGAATTTGGTGGCAGATATGACCAAAATTTTGA 8843
Db 1662 ACAAGACGATATTTATGAAAGCTGACGGAATTTGGTGGCAGATATGACCAAAATTTTGA 1721
QY 8844 TAACTCTGCTACTACAAATCCAAAGTGAATTTTACCAGTGTGCGAAGTCTCGA 8903
Db 1722 TAACTCTGCTACTACAAATCCAAAGTGAATTTTACCAGTGTGCGAAGTCTCGA 1781
QY 8904 ATCATTTCTTTGTACAGAAATTTGAAAGCTTCAAAGCTTAGCAGGTCTCTATACCAACAACACT 8963
Db 1782 ATCATTTCTTTGTACAGAAATTTGAAAGCTTCAAAGCTTAGCAGGTCTCTATACCAACAACACT 1841
QY 8964 GCAGTCTACAGCTTCTTAAAGTTCAGCGTGTAACTTAACATAAACAACAGCAAGATCT 9023
Db 1842 GCAGTCTACAGCTTCTTAAAGTTCAGCGTGTAACTTAACATAAACAACAGCAAGATCT 1901
QY 9024 GGTGTCTGACTATTTTAAATTAAGGACCATGTTTTTAGTCAGGCTATCTCTGCAAA 9083
Db 9083 GGTGTCTGACTATTTTAAATTAAGGACCATGTTTTTAGTCAGGCTATCTCTGCAAA 9083
```

Db 1902 GGTGTCTGAACATTTAAATTAAGGAGCCAGATGTTTTTACTCAGGCTATCCTTGACAA 1961
Qy 9084 GACTTGACCTAAACTTCGTTTTTATTTGGTCATACAGTCCCAATATATCTTGGCCAAAT 9143
Db 1962 GACTTGACCTAAACTTCGTTTTTATTTGGCCATACAGTCCCAATATATCTTGGCCAAAT 2021
Qy 9144 TTGTCCACGCAAGCAAGAAAAAGCAAGTCAACGACACACCATTTATCTTGTCAAGATCAGAT 9203
Db 2022 TTGTCCACGCAAGCAAGAAAAAGCAAGTCAACGACACACCATTTATCTTGTCAAGATCAGAT 2081
Qy 9204 GGTTTTACTATTGTGGCAGAGGAGAGAAAAACTTTGTTTAFTG 9245
Db 2082 GGTTTTACTATTGTGGCAGAGGAGAGAAAAACTTTGTTTAFTG 2123

RESULT 14
AAH09968/C
ID AAH09968 standard; cDNA; 556 BP.
XX
AC AAH09968;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (3'-primer) SEQ ID NO:6803.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
FN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.

XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 3; SEQ ID 6803; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 556 BP; 158 A; 106 C; 112 G; 175 T; 5 other;
Query Match 5.5%; Score 538.2; DB 22; Length 556;
Best Local Similarity 97.7%; Pred. No. 5.1e-97;
Matches 543; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 8690 GGCCTTTCCTTGAACAGTAGACCCCTAATGATGCCACCAATTTATGATGTTTAAAG 8749
Db 556 GGCCTTTCCTTGAACAGTAGACCCCTAATGATGCCACCAATTTATGATGTTTAAAG 497
Qy 8750 AACCTATGGACCTTGCACCATGGAAGAAAGATACAAAGACGATATTTGAAAACTGA 8809
Db 496 ACCNTATGGACCTTGCACCATGGAAGAAAGATGTCAAAGNCGATATTTGAAAACTGA 437
Qy 8810 CGGAATTTGTGGCAGATATGACCAAAATTTTGTAACTGTCTACTACAAATCCAAAGTG 8869
Db 436 CGGAATTTGTGGCAGATATGACCAAAATTTTGTAACTGTCTACTACAAATCCAAAGCG 377
Qy 8870 ACTCCCATTTTACCAGTGTGAGAAAGTTCTCGAATCATCTTTGTACAGAAATTCGAAAG 8929
Db 376 ACTCCCATTTTACCAGTGTGAGAAAGTTCTCGAATCATCTTTGTACAGAAATTCGAAAG 317
Qy 8930 GCTTCAAGCTAGCAGGTCTCATAACAACAACTGCAGTCTACAGCTTCTTAAAGTTCAG 8989
Db 316 GCTTCAAGCTAGCAGGTCTCATAACAACAACTGCAGTCTACAGCTTCTTAAAGTTCAG 257
Qy 8990 CGTGTAACTAACATATAAACAACAGCAAGAACTGTGTGTCTGAACATTTTAAATTAAG 9049
Db 256 CGTGTAACTAACATATAAACAACAGCAAGAACTGTGTGTCTGAACATTTTAAATTAAG 197
Qy 9050 GAGCCAGATGTTTTTACTCAGGCTATCCTGACAGACTTGCACCTAACTTCGTTTTTATT 9109
Db 196 GAGCCAGATGTTTTTACTCAGGCTATCCTGACAGACTTGCACCTAACTTCGTTTTTATT 137
Qy 9110 GGTCAATACAGTCCCAATATATTTTGGCCAAATTTTGTCCACGGACAGAAAAAGCAA 9169
Db 136 GGTCAATACAGTCCCAATATATTTTGGCCAAATTTTGTCCACGGACAGAAAAAGCAA 77
Qy 9170 AGTCAACGACACCATTTATCTTGTCAAGATCAGATGTTTACTATTTGTCAGAGCGAG 9229
Db 76 AGTCAACGACACCATTTATCTTGTCAAGATCAGATGTTTACTATTTGTCAGAGCGAG 17
Qy 9230 AAAACTTTGTTTATTG 9245
Db 16 AAAACTTTGTTTATTG 1

RESULT 15
AAZ17218
ID AAZ17218 standard; cDNA; 763 BP.
XX
AC AAZ17218;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:4689.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
FN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 09:12:14 ; Search time 12025 Seconds
(without alignments)
19938.766 Million cell updates/sec

Title: us-09-698-295-2
Perfect score: 9865
Sequence: 1 ggcagctgagtgccgc.....acaaaaaaaaaaaaaaaaa 9865

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	926	9.4	1086	12	BM458331
2	920.4	9.3	1131	12	BM802054
3	905.4	9.2	951	13	BU183985
4	892	9.0	1607	11	BC039881

5	857.8	8.7	897	13	BU187389
6	841	8.5	917	13	BQ716216
7	828	8.4	1046	12	BQ061478
8	823	8.3	867	13	BU169920
9	807.4	8.2	870	13	BQ718666
10	800.2	8.1	842	13	BU507358
11	799.4	8.1	883	13	BU176205
12	788.2	8.0	1200	11	AK082852
13	781.2	7.9	864	13	BQ717925
14	778	7.9	1050	12	BM451623
15	771.4	7.8	841	9	AU124130
16	732	7.4	1058	12	BM550180
17	727.2	7.4	786	13	BU739549
18	720.4	7.3	722	13	BX098782
19	713.8	7.2	905	10	BG164582
20	702.8	7.1	830	12	BI457269
21	693.6	7.0	711	14	CB305710
22	692.4	7.0	735	13	BQ575828
23	683.8	6.9	687	12	BQ021222
24	663.4	6.7	682	10	AW961200
25	662.4	6.7	683	2	HSM088021
26	661.4	6.7	746	12	BM675313
27	648.6	6.6	915	10	BG574526
28	638.2	6.5	968	13	BQ944857
29	636.8	6.5	648	12	BM312335
30	622.4	6.3	635	14	CB216388
31	620.4	6.3	894	14	CB181420
32	618.2	6.3	867	10	BE870101
33	615.2	6.2	695	10	BG393106
34	613	6.2	770	12	BI766147
35	612.4	6.2	673	12	BM992191
36	611	6.2	833	12	BI687006
37	609.8	6.2	643	13	BU674038
38	608.2	6.2	613	12	BQ001968
39	606	6.1	606	12	BM786451
40	601	6.1	677	9	AW390459
41	600.2	6.1	709	14	CB447515
42	595.2	6.0	741	9	AV704301
43	590.4	6.0	745	14	CB228541
44	589.8	6.0	718	12	BM996255
45	587.4	6.0	655	9	AV727211

ALIGNMENTS

RESULT 1
BM458331
LOCUS
DEFINITION AGNCOURT_6414182 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5496931
5', mRNA sequence.
ACCESSION BM458331
VERSION BM458331.1 GI:18507371
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1086)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: L1AM12126 row: d column: 20
High quality sequence stop: 702.

[illegible]


```

QY 8480 -----AATTTATATTGGCTGTGATCGGTGTGACAAATTGGTACCATGGCGCTGC 8529
Db 792 GATGACTCAAAATTTTATATTGGCTGTGATCGGTGTGACAAATTGGTACCATGGCGCTGC 851
QY 8530 GTTGGCATCTTGCAAAGTCAGGCGAGAGCTCATTTGATGAGTATCTCTGTCCACAGTCCAG 8589
Db 852 GTTGGCATCTTGCAAAGTCAGGCGAGAGCTCATTTGATGAGTATCTCTGTCCACAGTCCAG 911
QY 8590 TCAACAGAGATGCCATGATGCTTCACGCCACTAACAGAGAGGATTATGAGGGTTG 8649
Db 912 TCAACAGAGATGCCATGATGCTTCACGCCACTAACAGAGAGGATTATGAGGGTTG 971
QY 8650 AAGAGGGTCTCGTTTCCTTACAGGCCATAGATGGCTTGGCTTTCCTTGAACAGTA 8709
Db 972 AAGAGGGTCTCGTTTCCTTACAGGCCATAGATGGCTTGGCTTTCCTTGAACAGTA 1031
QY 8710 GACCTTAATGATGCACAGATATTATGGTGTATTATTAAGGAACCTATGACCTTGGCAC 8769
Db 1032 GACCTTAATGATGCACAGATATTATGGTGTATTATTAAGGAACCTATGACCTTGGCAC 1091
QY 8770 ATGGAAGAAGAGTACAAAGACGATATTATGAAAAGCTGACGGAATTTGTGGCAGATAG 8829
Db 1092 ATGGAAGAAGAGTACAAAGACGATATTATGAAAAGCTGACGGAATTTGTGGCAGATAG 1151
QY 8830 ACCAAATTTTGAATACACTGCTGTACTACATCCAAAGTACTCCCATTTTACAGTGT 8889
Db 1152 ACCAAATTTTGAATACACTGCTGTACTACATCCAAAGTACTCCCATTTTACAGTGT 1211
QY 8890 GCAGAAGTCTCGAATCATTTTGTACAGAAATTCAGAGGCTTCAAGCTAGCAGGT 8947
Db 1212 GCAGAAGTCTCGAATCATTTTGTACAGAAATTCAGAGGCTTCAAGCTAGCAGAT 1269

```

```

RESULT 5
LOCUS BU187389
DEFINITION AGENCOURT_7782656 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6137255
5', mRNA sequence.

```

```

ACCESSION BU187389
VERSION BU187389.1 GI:22701373
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 897)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13450 row: 1 column: 24
High quality sequence stop: 727.
Location/Qualifiers

```

```

FEATURES
Source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6137255"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

```

```

RESULT 6
LOCUS BU187389
DEFINITION AGENCOURT_7782656 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6137255
5', mRNA sequence.

```

```

ACCESSION BU187389
VERSION BU187389.1 GI:22701373
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 897)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13450 row: 1 column: 24
High quality sequence stop: 727.
Location/Qualifiers

```

```

FEATURES
Source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6137255"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

```

```

BASE COUNT 257 a 294 c 179 g 167 t
ORIGIN
Query Match 8.7%; Score 857.8; DB 13; Length 897;
Best Local Similarity 98.0%; Pred. No. 2.4e-108;
Matches 879; Conservative 0; Mismatches 17; Indels 1; Gaps 1;
QY 6378 AGGGCTCAGATTCGGCCCTGGTATGACCTGATGACCCAGTATTAGAACACCACTCCACAGTCAACACT 6437
Db 1 AGGGCTCAGATTCGGCCCTGGTATGACCTGATGACCCAGTATTAGAACACCACTCCACAGTCAACACT 60
QY 6438 AGGAAGGCAATTAATTCGAACACCTGTGATGGTACAGCCAGGTGCTCCTCAGCAAGTGAT 6497
Db 61 AGGAAGGCAATTAATTCGAACACCTGTGATGGTACAGCCAGGTGCTCCTCAGCAAGTGAT 120
QY 6498 GACTCAAAATCATCAGGGGGCAGCTGTCTCCACTGCACTGCTCCGCCCTCAACCGGTTTC 6557
Db 121 GACTCAAAATCATCAGGGGGCAGCTGTCTCCACTGCACTGCTCCGCCCTCAACCGGTTTC 180
QY 6558 CTCAACACCTGGGCGAGAAAGTTAACTTCAGCAAGCTCCACTTCAAAATATACAGTCTTC 6617
Db 181 CTCAACACCTGGGCGAGAAAGTTAACTTCAGCAAGCTCCACTTCAAAATATACAGTCTTC 240
QY 6618 AGCTCACAAACCCCTCGCCCCCAACAGGACAGTGAAGCTCACCATGCTCAACTTAC 6677
Db 241 AGCTCACAAACCCCTCGCCCCCAACAGGACAGTGAAGCTCACCATGCTCAACTTAC 300
QY 6678 TCAGTTAAACACAGGGGCGAGGTCGCAATCAAGGTTTGACAGTAGTAGTAATTCAGGACAAG 6737
Db 301 TCAGTTAAACACAGGGGCGAGGTCGCAATCAAGGTTTGACAGTAGTAGTAATTCAGGACAAG 360
QY 6738 TCAAACTACTGGAGCTTGCAGTTGATACCTCAAGGGGTGACTGTACTCCAGGCCAGG 6797
Db 361 TCAAACTACTGGAGCTTGCAGTTGATACCTCAAGGGGTGACTGTACTCCAGGCCAGG 420
QY 6798 CAGCAGCTTAATGCAAGCTGCAATGCCAATGTGTACTGTTTACCGGATTCCTCTTTACCCC 6857
Db 421 CCAGCAGCTTAATGCAAGCTGCAATGCCAATGTGTACTGTTTACCGGATTCCTCTTTACCCC 480
QY 6858 ATGGCAACAACAGCCACAGCCAGCCACACACACACACACACACACACACACACACAGC 6917
Db 481 ATGGCAACAACAGCCACAGCCACACACACACACACACACACACACACACACAGC 540
QY 6918 AGGTACAGGTGAACAAAGGCGAGTAGTAACCTGTACCCCGAGTGCAGGTACATCAAGACAA 6977
Db 541 AGGTACAGGTGAACAAAGGCGAGTAGTAACCTGTACCCCGAGTGCAGGTACATCAAGACAA 600
QY 6978 AACCTGCCACAGCTCAGTCATCAAGTGTGGTCCAGCAAAAGCCAGCCACAGACTGC 7037
Db 601 AACCTGCCACAGCTCAGTCATCAAGTGTGGTCCAGCAAAAGCCAGCCACAGACTGC 660
QY 7038 TCAGGCTTCAGCTCGGGCCCGAGCCCAACCCAGCCCGAGTCCCGAGCTCAGCTGAAGT 7097
Db 661 TCAGGCTTCAGCTCAGCCCGAGCCCAACCCAGCCCGAGTCCCGAGCTCAGCTGAAGT 720
QY 7098 TCAGACTAGCTGAAAGTTTCAGACCCCAAACTGTTTTCATCCCATGTCCTCTTCAAGC 7157
Db 721 TCAGACTAGCTGAAAGTTTCAGACCCCAAACTGTTTTCATCCCATGTCCTCTTCAAGC 780
QY 7158 ACNACCCCGAGGACAGTCTATCCAGCCCGAGTTCAGGACAGCTCAGCTCAAG 7217
Db 781 ACNACCCCGAGGACAGTCTATCCAGCCCGAGTTCAGGACAGCTCAGCTCAAG 840
QY 7218 TAATGCCAAGGACAGTCT-CTGTGTTGTTTCCAAAGTTCATCAGACAGCTCAAGTATC 7273
Db 841 TAATGCCAAGGACAGTCTCTCCCTGTTGTTGTTTCCAAAGTTCATCAGACACTTCGATAC 897

```

```

RESULT 6
LOCUS BU187389
DEFINITION AGENCOURT_8232760 Lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6191588 5', mRNA sequence.

```

```

BO716216 917 bp mRNA linear EST 16-JUL-2002
AGENCOURT_8232760 Lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6191588 5', mRNA sequence.

```

```

ACCESSION      B0716216
VERSION        B0716216.1  GI:21855113
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens

REFERENCE
AUTHORS        NIH-MGC http://mgc.nci.nih.gov/
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: ccapbs-r@mail.nih.gov
               Tissue Procurement: Dr. James R. Lupski
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Agencourt Bioscience Corporation
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: L1AM13592 row: d column: 21
               High quality sequence stop: 680.
               Location/Qualifiers
                 1. .917
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:6191588"
                   /sex="male"
                   /tissue_type="sympathetic trunk"
                   /dev_stage="adult, 16 yr"
                   /lab_host="DH10B"
                   /clone_lib="Lupski-sympathetic_trunk"
                   /note="vector: pCMV-SPORT6 (Life Technologies); Site_1:
                   NotI; Site_2: SalI; cDNA made by oligo-dT priming.
                   Directionally cloned using the following adaptors:
                   5'-TCGACCCACGGCTCCG-3' and
                   5'-GACTAGTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected >
                   1 kb for average insert length 1.9 kb. This is a primary
                   library, non-amplified. Library constructed by Life
                   Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
                   College of Medicine); available through Life
                   Technologies."
BASE COUNT     284 a 251 c 196 g 176 t 10 others
ORIGIN
Query Match.      8.5%; Score 841; DB 13; Length 917;
Best Local Similarity 98.5%; Pred. No. 4.8e-106;
Matches 867; Conservative 0; Mismatches 10; Indels 3; Gaps 2;

QY 5809 TCAAGTGCACATCGCGCCCAAGAGACACGAGAACGCCCAAGCAAACTGGCCCTGTATTATT 5868
DB      |||||||
QY 5869 GAAACCTGGGTAGCAGAGAAGAACTGGAATTTGGGAGATCAGGGCATTTGCTGAGAGA 5928
DB      |||||||
QY 5929 GTGGAGAAAGAAAGGCAACAGCAGTTGAGCAACAGGCTAAGAAACGACTGGAGCAGCAG 5988
DB      |||||||
QY 121  GTGGAGAAAGAAAGGCAACAGCAGTTGAGCAACAGGCTAAGAAACGACTGGAGCAGCAG 180
DB      |||||||
QY 5989 AAGCCGACAGTGATTTGCACATTTCCACTACTTTCCCAACAAGCAGTACACCCAGCACCATC 6048
DB      |||||||
QY 181  AAGCCGACAGTGATTTGCACATTTCCACTACTTTCCCAACAAGCAGTACACCCAGCACCATC 240
DB      |||||||
QY 6049 TCTCCAGCAGAGAAGTTATGGTGGCCCCCATAAGTGGCTCAGTTACAACTGGAAACCAA 6108
DB      |||||||
QY 241  TCTCCAGCAGAGAAGTTATGGTGGCCCCCATAAGTGGCTCAGTTACAACTGGAAACCAA 300
DB      |||||||
QY 6109 ATGGTACTAACTACTAAAGTTGGATCTCCAGCTCAGTAACTATCCCAACAAAGAAC 6168
DB      |||||||
QY 301  ATGGTACTAACTACTAAAGTTGGATCTCCAGCTCAGTAACTATCCCAACAAAGAAC 360
DB      |||||||

QY 6169 TTTTCATCAAAACCTTTTCTACATGGGTTAAGCAAGGCCAGTCAAATTCAGGCGTTGTCAA 6228
DB      |||||||
QY 361  TTTTCATCAAAACCTTTTCTACATGGGTTAAGCAAGGCCAGTCAAATTCAGGCGTTGTCAA 420
DB      |||||||
QY 6229 GTACAGCAGAAAGTCTCTGGGTATTCATTCATCAAGTACAGGTACAGTACAGCAAACTTT 5288
DB      |||||||
QY 421  GTACAGCAGAAAGTCTCTGGGTATTCATTCATCAAGTACAGGTACAGTACAGCAAACTTT 480
DB      |||||||
QY 6289 ACTTCATTTCCAGCCAGGACACGACAGTCAAAATTTAGCCCAATACCTCAGGCTCTGGA 6348
DB      |||||||
QY 481  ACTTCATTTCCAGCCAGGACACGACAGTCAAAATTTAGCCCAATACCTCAGGCTCTGGA 540
DB      |||||||
QY 6349 GGAACCCACAAGCAATTCACAAGTAATCACAGGCGCTCAGATTCGCCCTGGTATGACCGTG 6408
DB      |||||||
QY 541  GGAACCCACAAGCAATTCACAAGTAATCACAGGCGCTCAGATTCGCCCTGGTATGACCGTG 600
DB      |||||||
QY 6409 ATTAGAACCCACTCTCAACAGTCAACACTAGGAAAGCAATTTATTCGAACACCTGTGATG 6468
DB      |||||||
QY 601  ATTAGAACCCACTCTCAACAGTCAACACTAGGAAAGCAATTTATTCGAACACCTGTGATG 660
DB      |||||||
QY 6469 GTACAGCAGGTGCTCTCTCAGCAAGTGATGACTCAAAATCATCAGGGGGCAGCGTGTCTCC 6528
DB      |||||||
QY 661  GTACAGCAGGTGCTCTCTCAGCAAGTGATGACTCAAAATCATCAGGGGGCAGCGTGTCTCC 720
DB      |||||||
QY 6529 ACTGCAGTCTCCGCCCTCAACAGGTTTCTCAACACCTGGGCGAGAAAAGCTTAACCTCA 6588
DB      |||||||
QY 721  ACTGCAGTCTCCGCCCTCAACAGGTTTCTCCCTCACACCTGNGCAGAAAAGCTTAACCTCA 780
DB      |||||||
QY 6589 GCAAGCTCCACTTCAATATACAGTCTTCAGCTCACAACCCCTCGCCCCCAACAAGGA 6648
DB      |||||||
QY 781  GCAAGCTCCACTTCAATATACAGTCTTCAGCTCACAACCCCTCGCCCCCAACAAGGA 840
DB      |||||||
QY 6649 C-AAGTGAAGCTCACCATGG--CTCAACTTACTCAGTTAA 6685
DB      |||||||
QY 841  C-NNAGTGAAGCTCACCATGGGNGCTCANCCTTACTCAGTTAA 880
DB      |||||||

RESULT 7
BQ061478
LOCUS     BQ061478
DEFINITION 5', mRNA sequence.
ACCESSION BQ061478
VERSION   BQ061478.1  GI:19885571
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1046)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: ccapbs-r@mail.nih.gov
          Tissue Procurement: Lou Staudt
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LICM2085 row: 9 column: 09
          High quality sequence stop: 674.
          Location/Qualifiers
            1. .1046
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:5920928"
              /tissue_type="lymphoma, cell line"
              /lab_host="DH10B (phage-resistant)"
FEATURES
source

```

/clone_lib="NIH_MGC_99"
/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Birectionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT 368 a 226 c 209 g 243 t

ORIGIN

Query Match 8.4%; Score 828; DB 12; Length 1046;

Best Local Similarity 96.9%; Pred. No. 2.9e-104;

Matches 899; Conservative 0; Mismatches 20; Indels 9; Gaps 5;

QY 4507 AGAGTAGTAAGTGGTAATGTTGAACCAAGGTTAATAATATAATAATAATAATATCCCTGAG 4566

DB 1 AGAGTAGTAAGTGGTAATGTTGAACCAAGGTTAATAATATAATAATAATATCCCTGAG 60

QY 4567 AATGATATTAAATCATCTGACTGTTAAAGAACTGCTATAAGGCCATTCAATTAATGGTGTAT 4626

DB 61 AATGATATTAAATCATCTGACTGTTAAAGAACTGCTATAAGGCCATTCAATTAATGGTGTAT 120

QY 4627 GTCATCTGGAAGATTTTAATGAAGAAACAGCTCCGAAACAAATATCGATTGCTGTGAGT 4686

DB 121 CTCATCATGGAGATTTTAATGAAGAAACAGCTCCGAAACAAATATCGATTGCTGTGAGT 180

QY 4687 TCTTCAGATGCTGAAGTAACTTACCGAGATAGCCTTGAGACCCCTGCCATCAACCAAGAG 4746

DB 181 TCTTCAGATGCTGAAGTAACTTACCGAGATAGCCTTGAGACCCCTGCCATCAACCAAGAG 240

QY 4747 TCTGACGTACACAGACACACACCCCTCAGCATCTTGTCAGAAAGCAATTCAGTTAAT 4806

DB 241 TCTGACGTACACAGACACACACCCCTCAGCATCTTGTCAGAAAGCAATTCAGTTAAT 300

QY 4807 CAGGTAGAAGATATGGAATAGAAACCTCAGAAGTTAAGAAGTTACTATCACTATT 4866

DB 301 CAGGTAGAAGATATGGAATAGAAACCTCAGAAGTTAAGAAGTTACTATCACTATT 360

QY 4867 ACTTCTGAAGAGGAATCTAATCTCAGTATGACTTTATTGATGAATGGTGTGCGCCATC 4926

DB 361 ACTTCTGAAGAGGAATCTAATCTCAGTATGACTTTATTGATGAATGGTGTGCGCCATC 420

QY 4927 AACAAAAATGAAATGTCAATGGAGAATCTAAAGAAAAACCGTCATCACAGAACTACC 4986

DB 421 AACAAAAATGAAATGTCAATGGAGAATCTAAAGAAAAACCGTCATCACAGAACTACC 480

QY 4987 ACGATGACCTCCAGTGGCCACAGAAATCAAAAATGTGATCAAGGTAGAAAAAGGCGAT 5046

DB 481 ACGATGACCTCCAGTGGCCACAGAAATCAAAAATGTGATCAAGGTAGAAAAAGGCGAT 540

QY 5047 AAGCAAACTGGTGTCTTCACACAGAAATGTGCAGAAATCCACTGTCCACACCACTACT 5106

DB 541 AAGCAAACTGGTGTCTTCACACAGAAATGTGCAGAAATCCACTGTCCACACCACTACT 600

QY 5107 ACAACAGTGACCAAGCTTTCCACACCCCTCCACAGCGCGAGTGTGGACATCATCTCTGTA 5166

DB 601 ACAACAGTGACCAAGCTTTCCACACCCCTCCACAGCGCGAGTGTGGACATCATCTCTGTA 660

QY 5167 AAGGAGCAGAGAAACCGTGTACACAGCAGTGACAGACTCCCTGACACACACCGGGA 5226

DB 661 AAGGAGCAGAGAAACCGTGTACACAGCAGTGACAGACTCCCTGACACACACCGGGA 720

QY 5227 GGCACACTGGTTACATCTATGACTGTGAGCAAGAGTATCCACAGCAGAGAAAGTGAAA 5286

DB 721 GGCACACTGGTTACATCTATGACTGTGAGCAAGAGTATCCACAGCAGAGAAAGTGAAA 780

QY 5287 CTGATGAATTTTCAAGA-CCAAAAGAGACT-CGTTTCAGGTACAGCTCTGCGCAT-CCTAT 5343

DB 781 CTGATGAATTTTCAAGACCCCAAGAGACTCCCGTTTCAGGTACAGCTCTGCGCATCCCTAT 840

QY 5344 AGAAAAATTTGTACCAAG-----AGCACCAAGAAAGACCATTTTGTTCCTCTAA--TGAT 5397

DB 841 AGAAAAATTTGTACCAAGGAGCAGCAAGGAAGAGCATTTTGTTCCTCTAAATGATG 900

QY 5398 GACTTAAAAAAGCTTGGCCGCAAAAGGAG 5425

DB 901 GACTTAAAAAAGCTTGGCCGCAAAAGGG 928

RESULT 8

BU169920'

LOCUS BU169920 867 bp mRNA linear EST 04-SEP-2002

DEFINITION AGENCOURT_7938595 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6012389

5', mRNA sequence.

ACCESSION BU169920

VERSION BU169920.1 GI:22683904

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 867)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM13203 row: j column: 06

High quality sequence stop: 690.

Location/Qualifiers

1..867

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6012389"

/tissue_type="embryonal carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_92"

/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 2.5 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC Library."

BASE COUNT 303 a 170 c 195 g 196 t 3 others

ORIGIN

Query Match 8.3%; Score 823; DB 13; Length 867;

Best Local Similarity 97.8%; Pred. No. 1.5e-103;

Matches 843; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 1970 GTAACAGCAGCAGTGAACCTAAATTTCTCCAGAGTGAATCTCTAAGGCAGCTGATGATC 2029

DB 1 GTAACAGCAGCAGTGAACCTAAATTTCTCCAGAGTGAATCTCTAAGGCAGCTGATGATC 60

QY 2030 CTGAAAATGGGAAAGAGAAATCTATACACCTGTCTCTATTTCAGGAAGAGATAGTAGTG 2089

DB 61 CTGAAAATGGGAAAGAGAAATCTATACACCTGTCTCTATTTCAGGAAGAGATAGTAGTG 120

QY 2090 ATTTTCATCTCGAGAGTCCACCGGGAGCTAAAGTGAATCTCTCTCGAGCTGGAAAGGAG 2149

DB 121 ATTTTCATCTCGAGAGTCCACCGGGAGCTAAAGTGAATCTCTCTCGAGCTGGAAAGGAG 180

QY 2150 CATCTGGCTCACTCGAATCATCACAGATTCGGGAATCCAGATAGCAAACTTAGTCAGC 2209

DB 181 CATCTGGCTCACTCGAATCATCACAGATTCGGGAATCCAGATAGCAAACTTAGTCAGC 240

[illegible]

```

Db      661  TATGAGGGGTTGAAGAGGTGCTCCGTTCTTACAGCCCAATAGATGGCCCTGGCCCTTT 720
QY      8697  CTTGAACAGTAGACCCCTAATGATGCACAGATTATTATGGTGTATTAAAGAACCTAT 8756
Db      721  CCTTGAACAGTAGACCCCTAATGATGCACAGATTATTATGGGCTATTAAAGAACCTAT 780
QY      8757  GGACCTTCCACCATGGGAAGAGAGTACAAGACGATATTATGAAGAGCTG--ACGGAA 8814
Db      781  GGACCTTCCACCATGGGAAGAGAGTACAAGACGATATTATTGAAGAGCTGGACGGAA 840
QY      8815  TTGTGTGGCAGATATGACCAAAATTTT 8841
Db      841  TTGGGGGAGATATGACCAAAATTTT 867

RESULT 10
BU507358
LOCUS   BU507358      842 bp      mRNA      linear      EST 12-SEP-2002
DEFINITION AGENCOURT_10309897 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6501281
5', mRNA sequence.
ACCESSION BU507358
VERSION   BU507358.1 GI:22813591
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 842)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-re@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM14055 row: h column: 18
          High quality sequence stop: 751.
FEATURES             source
     source
     1..842
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="IMAGE:6501281"
     /tissue_type="leiomyosarcoma"
     /lab_host="DH10B (phage-resistant)"
     /clone_lib="NIH_MGC_71"
     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
     Average insert size 2.1 kb."
BASE COUNT  257 a 246 c 171 g 167 t 1 others
ORIGIN
Query Match      8.1%; Score 800.2; DB 13; Length 842;
Best Local Similarity 99.4%; Pred. No. 2e-100;
Matches 813; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY      6129  TGGATCTCCAGTACAGTAACATCCAAACAAAGAACTTTTCATCAAACTTTTGCTAC 6188
Db      1    TGGATCTCCAGTACAGTAACATCCAAACAAAGAACTTTTCATCAAACTTTTGCTAC 60
QY      6189  ATGGGTTAAGCAAGCCAGTCAATTCAGGGGTGTTCAAGTACAGCAAGTCTCTGG 6248
Db      61    ATGGGTTAAGCAAGCCAGTCAATTCAGGGGTGTTCAAGTACAGCAAGTCTCTGG 120
QY      6249  TATCATTCATCAAGTACAGTACCAGTCAGCAAAACCTTTACTTTCATTCAGCCCGAGAC 6308
Db      121  TATCATTCATCAAGTACAGTACCAGTCAGCAAAACCTTTACTTTCATTCAGCCCGAGAC 180

```

FEATURES

```

QY      6309  AGCAACAGTCAAAATTAGGCCCAATACCTCAGGCTCTGGAGGAACACCAAGCAATTACA 6368
Db      181  AGCAACAGTCAAAATTAGGCCCAATACCTCAGGCTCTGGAGGAACACCAAGCAATTACA 240
QY      6369  AGTAATACAGGGCCTCAGATTGCGCCCTGCTGATGACCGTGATTAGAACACCACTCCAACA 6428
Db      241  AGTAATACAGGGCCTCAGATTGCGCCCTGCTGATGACCGTGATTAGAACACCACTCCAACA 300
QY      6429  GTCAACACTAGGAAAGGCAATTTATTGCAACACCTGTGATGGTACAGCCAGGTGCTCTCA 6488
Db      301  GTCAACACTAGGAAAGGCAATTTATTGCAACACCTGTGATGGTACAGCCAGGTGCTCTCA 360
QY      6489  GCAAGTGATGACTCAAAATCATCAGGGGGCAGCCTGTCTCCACTGCGAGTCTCCGCTTAA 6548
Db      361  GCAAGTGATGACTCAAAATCATCAGGGGGCAGCCTGTCTCCACTGCGAGTCTCCGCTTAA 420
QY      6549  CACGGTTTCTTCAACACACCTGGGCAGAAAGCTTAACCTTCAGCAACGTCACATTCAAATAT 5608
Db      421  CACGGTTTCTTCAACACACCTGGGCAGAAAGCTTAACCTTCAGCAACGTCACATTCAAATAT 480
QY      6609  ACAGTCTTTCAGCTCACAACACCCCTCGCCCCCAACAAAGGACAAGTGAAGCTCACCATGCG 6668
Db      481  ACAGTCTTTCAGCTCACAACACCCCTCGCCCCCAACAAAGGACAAGTGAAGCTCACCATGCG 540
QY      6669  TCAACTTACTAGTTAAACACAGGGCCAGCGTGGCAATCAAGCTTTGACAGTAGTAGTAATCA 6728
Db      541  TCAACTTACTAGTTAAACACAGGGCCAGCGTGGCAATCAAGCTTTGACAGTAGTAGTAATCA 600
QY      6729  AGGACAAGGTCAAACTACTGGACAGTTGCGAGTTGATACCTCAAGGGGTGACTGTACTCCG 6788
Db      601  AGGACAAGGTCAAACTACTGGACAGTTGCGAGTTGATACCTCAAGGGGTGACTGTACTCCG 660
QY      6789  AGGCCAGCGCCAGCAGCTAATGCAAGCTGCAATGCCAATGCTACTGTTTCACGGATTCTCT 6848
Db      661  AGGCCAGCGCCAGCAGCTAATGCAAGCTGCAATGCCAATGCTACTGTTTCACGGATTCTCT 720
QY      6849  CTTTACCCCTATGGCAACAACAGCCACACAGCCAGCCAGCCACACACACCTGTTTCCAC 6908
Db      721  CTTTACCCCTATGGCAACAACAGCCACACAGCCAGCCAGCCAGCCACACACCTGTTTCCAC 780
QY      6909  GACAGCAGCAGGTACA-GGTGAACAAGCGACAGATAA 6945
Db      781  GACAGCAGCAGGTACAGGGTGAACAAGCGACAGATAA 818

RESULT 11
BU176205
LOCUS   BU176205      883 bp      mRNA      linear      EST 04-SEP-2002
DEFINITION AGENCOURT_7827724 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6139068
5', mRNA sequence.
ACCESSION BU176205
VERSION   BU176205.1 GI:22690189
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 883)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-re@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM13455 row: h column: 13
          High quality sequence stop: 662.
          Location/Qualifiers

```

source	1. .883	/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="taxon:9606"
		/clone="IMAGE:6139068"
		/tissue_type="retinoblastoma"
		/lab_host="DH10B (phage-resistant)"
		/clone_lib="NIH_MGC_67"
		/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
		Average insert size 1.75 kb. Library constructed by Life Technologies."
BASE COUNT	289 a 199 c 202 g 191 t	2 others
ORIGIN		
Query Match	8.1%;	Score 799.4; DB 13; Length 883;
Best Local Similarity	98.0%;	Pred. No. 2.6e-100;
Matches 862;	Conservative 0;	Mismatches 12; Indels 6; Gaps 5;
QY	5027	TCAAGGTAGAAAAGCGGATAACAAACTCTGGTTCTTCCACAGAAAAATTGTCRAAAT 5086
Db	1	TCAAGGTAGAAAAGCGGATAACAAACTCTGGTTCTTCCACAGAAAAATTGTCRAAAT 60
QY	5087	CCACTGTCAACAACCACCTACACAGTGACCAAGCTTTCCACACCCCTCCACAGGCGCA 5146
Db	61	CCACTGTCAACAACCACCTACACAGTGACCAAGCTTTCCACACCCCTCCACAGGCGCA 120
QY	5147	GTGTGGACATCATCTCTGTTAAAGAGCAGAGCAAAACCGTGGTCCACACACAGTGACAG 5206
Db	121	GTGTGGACATCATCTCTGTTAAAGAGCAGAGCAAAACCGTGGTCCACACACAGTGACAG 180
QY	5207	ACTCCCTGACCAACCGGGAGGCACACTGGTTACATCTATGACTGTGAGCAAAAGAGTATT 5266
Db	181	ACTCCCTGACCAACCGGGAGGCACACTGGTTACATCTATGACTGTGAGCAAAAGAGTATT 240
QY	5267	CCACACGAGACAAAGTGAATGATGAAATTTTCAAGACCAAAGAAGACTCGTTTCAGGTA 5326
Db	241	CCACACGAGACAAAGTGAATGATGAAATTTTCAAGACCAAAGAAGACTCGTTTCAGGTA 300
QY	5327	CAGCTCGCCATCCTATAGAAAAATTTGTTACCAAGAGCACCAAGAAGACATTTTTCTGTT 5386
Db	301	CAGCTCGCCATCCTATAGAAAAATTTGTTACCAAGAGCACCAAGAAGACATTTTTCTGTT 360
QY	5387	TGCCTAATGATGACTTAAAAAAGTTGCCCGAAAAAGGAGGAATCCGAGAGGTCCCTTATT 5446
Db	361	TGCCTAATGATGACTTAAAAAAGTTGCCCGAAAAAGGAGGAATCCGAGAGGTCCCTTATT 420
QY	5447	TTAATTACAATGAAAACCTGCTTTGGATATATGGCCATATCCCTTCCTTAGACCCGACCT 5506
Db	421	TTAATTACAATGAAAACCTGCTTTGGATATATGGCCATATCCCTTCCTTAGACCCGACCT 480
QY	5507	TTGSCATCACTTTGGAGTATAGACTTCAGACAGTAAAGTCTTAGCTGGAGTGAGCCTGA 5566
Db	481	TTGSCATCACTTTGGAGTATAGACTTCAGACAGTAAAGTCTTAGCTGGAGTGAGCCTGA 540
QY	5567	TGTTACGGTTACTGTGGGCAAGTTTGAGATGGGATGATATGGGGCCCAAGGTTCCCTCCAG 5626
Db	541	TGTTACGGTTACTGTGGGCAAGTTTGAGATGGGATGATATGGGGCCCAAGGTTCCCTCCAG 600
QY	5627	GAGGAGGAGGTACACGGACAGAACATCCGAAACTGAAATCACAACAACAGAAATAATTA 5686
Db	601	GAGGAGGAGGTACACGGACAGAACATCCGAAACTGAAATCACAACAACAGAAATAATTA 660
QY	5687	AGAGGAGAGATGTTGGTCCCTTATGGCATTCGATTTGGAATATGTGATCAGGAAATCATTT 5746
Db	661	AGAGGAGAGATGTTGGTCCCTTATGGCATTCGATTTGGAATATGTGATCAGGAAATCATTT 720
QY	5747	GTCCC - ATTGGAGTTCCAGAAACACCAAAAGAACCGCTACACCTCAGAGGAAA - GCCT 5804
Db	721	GTCCCNATTGGAGTTCCAGAAACACCAAAAGAACCGCTACACCTCAGAGGAAAGGCGCT 780
QY	5805	TCGATCAAGTGCATCGCGGCCAAG - AGACACAGAAAGCCCCAACCAACTGG - CCCTGTT 5862

TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs			
JOURNAL REFERENCE	Nature 420, 563-573 (2002)			
AUTHORS	6 (bases 1 to 1200) Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tanaka,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tagami,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)			
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/. Location/Qualifiers 1. .1200 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:C330043E15" /db_xref="taxon:10090" /clone="C330043E15" /cell_type="ES cells" /clone_lib="RIKEN full-length enriched mouse cDNA library" 37. .336 /note="unnamed protein product; putative unclassifiable" /codon_start=1 /protein_id="BAC38653.1" /db_xref="GI:26350027" /translation="MAWPFLEVPDNDADPDYGVIKPEMDLATMEERIKRYEKLTE FVADMTKIFDNCRYNRPDPFYQCAEVLESFFVQKLGFKASRSHNNKLQSTAP" 1176. .1181 polyA_signal /note="putative" 1200 polyA_site /note="putative" BASE COUNT 371 a 241 c 243 g 345 t ORIGIN			
Query Match	8.0%; Score 788.2; DB 11; Length 1200;			
Best Local Similarity	84.9%; Pred. No. 8.2e-99;			
Matches 1031; Conservative	0; Mismatches 153; Indels 31; Gaps 12;			
QY	8647	TTGAAGAGGCTGCTCGCTTCTTACAGGCCCATAGATGCGCTTTCCTTGAACCA	8706	
Db	1	TTGAAGAGGCTGCTCGCTCCTTACAGGCCCATAGATGCGCTTTCCTTGAACCG	60	
QY	8707	GTAGACCCCAATGATGACACAGATTTATTATGCTGTTATTAAAGAACCTATGACCTTGGC	8766	
Db	61	GTAGACCCCAATGATGACACCGATTATTACGCTGTTATTAAAGCAATGGACCTTGGC	120	
QY	8767	ACCATGGAAGAAGAGTACAAAGAGATATTATGAAAGCTGACGGAATTTGTGGCAGAT	8826	
Db	121	ACCATGGAAGAAGATACAAAACGGTATTATGAAAGCTGACAGAGTTCTGTGGCAGAT	180	
QY	8827	ATGACCAAAATTTTGTGATAACTGTGCTTACTACAATCCAAAGTGACTCCCAATTTTACCAG	8886	
Db	181	ATGACCAAAATTTTGTGATAACTGCGCTTATTACAATCCCGTGACACCCCTTTTACCAG	240	
QY	8887	TGTGCAAGAGTCTCCAAATCATTTTGTGTACAGAAATTCGAAAGGCTTCAAGAGCTACGAG	8946	
Db	241	TGTGCAAGAGTCTTCTGAATCATTTTGTGTACAGAAACTAAAGGATTTCAAGGCCACGAGG	300	
QY	8947	TCTCATAAACAACAATTCGAGCTGTACAGCTTCTTAAAGTTCTACGCGTGTAAACCTAACATA	9006	
Db	301	TCTCATAAACAACAAGCTGCAATCTACAGCTCTCTTAAGACTCAGCGTGTCTGTACACCTAAG	360	
QY	9007	AAACACAGCAAGAAATCTGGTTGTCTGAACTATTATTTAAATTAAGGACCGCAGATGTTTTTAG	9066	
Db	361	CTAGACAGCAAGTCTGGCGTCTGGAACCTA-TTTAAACTAAAGCCGCGAGATATTTTTCAG	419	
QY	9067	TCAGGCTATCCCTGACAAGACTTGACCTAAACTTCGTTTTTATTGGTCATACAAAGTCCAAT	9126	
Db	420	TCAGGCTTTCTTGACAA-----GACCGTAACTTCGTTTATTGGTCACAAAGTCCAGT	474	
QY	9127	TATATTCTTTGGGCAATTTTGTCCACGGCAAGAAAGCAAAAGTCAACGACACCATTA	9186	
Db	475	TGTATTCTTTGGCCAAATTTGTCTCCACGGCAAAAGCAAAAGTCAACGGCACCGCTTG	534	
QY	9187	TCTTGTCAAGATCAGATGGTTTACTATTGTGCGCAAGCGGAGAAAAACITTTGTTTATTGA	9246	
Db	535	TCTTGTGCGAGCAATGGCTTTACTATTGTGCGCAAGCAGGAAACTTTGTTTATTTGA	594	
QY	9247	AAAAAAGAGAAAGAAAG-----CAAGAAAAAAGATACTATGGGGTCAAGT	9294	
Db	595	AAAAAAGAGAAAGAAAGCAAGAAAAAAGAAAAAGATACTATGGGTCAGT	654	
QY	9295	GTAACCTCCATGGAAATGCCACGCTCTCTTTCAGTGAAGAAGCTGGTTAGAGTCTCACA	9354	
Db	655	GTAACCT-CCGGGGGATGCCACATCTCTCTCAGTGAAGAAGCT-GTTTAGAGTCTCAGA	712	
QY	9355	GAAACCTTTTGACTGTATTATTATTATTGTTGC--AAAAAGAGCGCTTTTATTGCTGCC	9412	
Db	713	AAACCTTTTGGACTGTATTATTATTGTTGCAAAAAAAGAGCGCTTTTATTGCTGCC	771	
QY	9413	CTCATTTGTCAGCTAAGTATTTTCTTATA-AAATCCAGCCCGCTTACATATAATCA-	9470	
Db	772	CTCATTTGTCAGCTAATTTATTTTCTTATACAAATCCAGCCCTGGTCCGTGTATCCAC	831	
QY	9471	TCTGTATCTTATCATGATTCCTGTAGGTAAAGTACAAAGACGACCTCTAGATGCTTTTC	9530	
Db	832	CCTGTATCTCTATCATGATTCCTGTAGGT-GAAGTACAAAGCGGACCTCTAGTGTCTTTTC	890	
QY	9531	TTTCTATGAAGAGGCTGTATGTACACATGTGCACACACACAACTGGGAATCAACAA	9590	
Db	891	TTTCTATGAAGAGGAGCTGTATGTACACATGTG---CACACAACTGGGAATCAACAA	946	
QY	9591	TGAGTTTATTGTTTCATGGTAGATTAAATTAAGCTTGCATAAAGTTGGGCTAAGTGGTC	9650	
Db	947	TGAGTTTATTGTTTCACGGGAGACACAAGTGAAGCTTCGGGAAAGTTGGGCTAAGTGGTC	1006	
QY	9651	CTTGGGCTACGACTCTGTTGCCCTTGAATATATACAGTACAAATTTGTCATTTACTCTGCAC	9710	
Db	1007	CTGGACTAACAGACTTCGTTGCCCTTGAATACACACAGTACAAATTTGTCATTTACTCTGCAC	1066	
QY	9711	CAGGCTAAAGTGAGTAAAAATCTATTGTAAGGATCTTGTGTTGTAACATTTTGTGCAGATTC	9770	
Db	1067	CAGGCTAAA-TGAGCAAAATCTATTGTAAGGATCTTGTGTTGTAACATTTGTCAGATTC	1125	
QY	9771	TAATTTTTTCTTTTGTATTAAAAATCAACTATGGATGTATATGAACAATAAATAATGA	9830	
Db	1126	AATTTTTTCTTTTGTATTAAAAATCAACTATGGATGTATACGAATTCGAAATAAATGA	1185	
QY	9831	GATAAATTTTCTCCC	9845	
Db	1186	GATCATTTTCTCTC	1200	

```

RESULT 13
BQ717925
LOCUS
DEFINITION BQ717925 864 bp mRNA linear EST 16-JUL-2002
AGENCOURT 8241359 Lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6187125 5', mRNA sequence.

ACCESSION BQ717925
VERSION BQ717925.1 GI:21856822
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 864)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13580 row: j column: 22
High quality sequence stop: 628.
FEATURES
source
Location/Qualifiers
1..864
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6187125"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/clone_lib="Lupski_sympathetic_trunk"
/vector="pCMV-SPORT6 (Life Technologies); Site_1:
Noti; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCG-3' and
5'-GACAGTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."
BASE COUNT 254 a 298 c 154 g 157 t 1 others
ORIGIN
Query Match 7.9%; Score 781.2; DB 13; Length 864;
Best Local Similarity 98.0%; Pred. No. 8.3e-98;
Matches 823; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

QY 6546 TAACACGGTTTCCTCAACACCTGGGAGAGAAAGCTTAAGTTCAGACAGCTCCACTTCAA 6605
Db
1 TAACACGGTTTCCTCAACACCTGGGAGAGAAAGCTTAAGTTCAGACAGCTCCACTTCAA 60

QY 6606 TATACAGTTCAGCTCAGACACCTCGCCGCCCAAGCAAGCAAGTGAAGTCAAGT 6665
Db
61 TATACAGTTCAGCTCAGACACCTCGCCGCCCAAGCAAGCAAGTGAAGTCAAGT 120

QY 6666 GGCTCAACTTACTCAGTTTAACACAGGGCGGCGGCAATCAAGTTTGACAGTAGTAAT 6725
Db
121 GGCTCAACTTACTCAGTTTAACACAGGGCGGCGGCAATCAAGTTTGACAGTAGTAAT 180

QY 6726 TCAAGGACAAGTCAAACTACTGGACAGTTGCAGTTGATACCTCAAGGGGTGACTGTACT 6785
Db
181 TCAAGGACAAGTCAAACTACTGGACAGTTGCAGTTGATACCTCAAGGGGTGACTGTACT 240

QY 6786 CCCAGGCCAGGCCAGCAGCTAATGCAAGCTGCAATGCAATGGTACTGTTTCAGCGGATT 6845

```

/lab_host="DH10B (phage-resistant)"									
/clone_lib="NIH_MGC_67"									
(note: eye; Vector: pCMV-SPORT6; Site_1: NotI;									
Site_2: SalI; Cloned unidirectionally. Primer: Oligo									
Average insert size 1.75 kb. Library constructed by Life									
Technologies."									
BASE COUNT	341 a	189 c	270 g	248 t	2	others			
ORIGIN									
Query Match	7.9%; Score 778; DB 12; Length 1050;								
Best Local Similarity	92.1%; Pred. No. 2.e-97;								
Matches	877;	Conservative	0;	Mismatches	62;	Indels	13;	Gaps	5;
Qy	648	CTTTTGTGCAGCTCTGTGTGAGCCAAAGCAGTGCACACTCATGSCAGAGATGCATGTTGT	707						
Db	13	CTTTTGTGCAGCTCTGTGTGAGCCAAAGCAGTGCACACTCATGSCAGAGATGCATGTTGT	72						
Qy	708	GCITTTGAAAGCAGTTCCTGGTGNAGAACACTTCCAATACTACCTTTGGACCTGCTGA	767						
Db	73	GCTTTTGAAGCAGTTCCTGGTGAAGAACACTTCCAATACTACCTTTGGACCTGCTGA	132						
Qy	768	TCTGAAAGATAGCGTTAAATCCACACTGATTTTCATAGATGGATGACGTGGCCAGAGGT	827						
Db	133	TCTGAAAGATAGCGTTAAATCCACACTGATTTTCATAGATGGATGACGTGGCCAGAGGT	192						
Qy	828	GCTCGGGTGTACTGTGAGAGTGATNAGAGTACCATCAGCTTCTCCTTACCAAGAGGC	887						
Db	193	GCTCGGGTGTACTGTGAGAGTGATNAGAGTACCATCAGCTTCTCCTTACCAAGAGGC	252						
Qy	888	AGAGGACTACCCATATGCGCAGTAGAGAACAGATCAAAAGTCTACAGTTTCTAGTCGA	947						
Db	253	AGAGGACTACCCATATGCGCAGTAGAGAACAGATCAAAAGTCTACAGTTTCTAGTCGA	312						
Qy	948	TCAGTTCTTTACACAAATATTCCTCAGAGGAATTCATGCTCAAGGGGTGATACAGTA	1007						
Db	313	TCAGTTCTTTACACAAATATTCCTCAGAGGAATTCATGCTCAAGGGGTGATACAGTA	372						
Qy	1008	TGATGACCATTTAGGGTTTGTACAAACTTTGGGGATTTCCTTGTCTGAGACATGTTTC	1067						
Db	373	TGATGACCATTTAGGGTTTGTACAAACTTTGGGGATTTCCTTGTCTGAGACATGTTTC	432						
Qy	1068	AGCAGTATACCATTTGGAATGTGTGAAGCACCTCTTGAGAGAGTGCCAGAGACGAGTG	1127						
Db	433	AGCAGTATACCATTTGGAATGTGTGAAGCACCTCTTGAGAGAGTGCCAGAGACGAGTG	492						
Qy	1128	GCAGTGTGAAGTCTGTGTAGCACACAGGTGCCTGGTGCACGTACGTCTTCTGTAAT	1187						
Db	493	GCAGTGTGAAGTCTGTGTAGCACACAGGTGCCTGGTGCACGTACGTCTTCTGTAAT	552						
Qy	1188	CCAAAAAATAAACCATATATTCGCATGAACCTATTGGATATGATAGAAGTCGGAGCAA	1247						
Db	553	CCAAAAAATAAACCATATATTCGCATGAACCTATTGGATATGATAGAAGTCGGAGCAA	612						
Qy	1248	ATACTGGTCTTGAACCGAAGACTCATATAGAGAGATACAGAAATGAAATGAAAA	1307						
Db	613	ATACTGGTCTTGAACCGAAGACTCATATAGAGAGATACAGAAATGAAATGAAAA	672						
Qy	1308	GAAAAATTGTATTACAGCAAAAGTCCAACTTGCAGAAATTAATTGACTGCTAGACAA	1367						
Db	673	GAAAAATTGTATTACAGCAAAAGTCCAACTTGCAGAAATTAATTGACTGCTAGACAA	732						
Qy	1368	AGATTATT- GGGAGCAGAACTCTGCAAAATCTTAGAAGAAATGCGTGAAGAAA- - - TC	1422						
Db	733	AGATTATTGGGAGCAGAACTCTGCAAAATCTTAGAAGAAATGCGTGAAGAAATCCAC	792						
Qy	1423	CACGCACATGGACATACCTGAGAGCCTGACCAATAAGG- - - CTCGGGCGAGTAACAAA	1479						
Db	793	CCGACACATGGNACATAACTGAAGACCTGACCAATAAGGCTCGGGGGCAGTAACAAA	852						
Qy	1480	TCC- TTTTCTGGGCGAGC- - - -TAACTGAAGAAATTTTGGAAATCCATAAGACCCAAAAGG	1534						
Db	853	TCCTTTTCTGGGCGAGCCTTAATGGAAAGAAAAATTTTGGAAATCCCTTAAGAGACCAAAAG	912						

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 09:12:14 ; Search time 404 Seconds
(without alignments)
10777.833 Million cell updates/sec

Title: US-09-698-295-2
Perfect score: 9865
Sequence: 1 ggcgcgctgagtggtgccc.....acaaaaaaaaaaaaaaaaaa 9865

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2.6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	814.6	8.3	1032	4	US-09-257-179-21
C 2	101.8	1.0	1926	4	US-09-249-585A-4
C 3	101.8	1.0	1931	2	US-09-130-114-2
C 4	93.2	0.9	7218	1	US-08-232-463-14
C 5	89	0.9	390	3	US-09-197-649-7
C 6	85.2	0.9	3489	2	US-08-728-323A-1
C 7	85.2	0.9	3489	4	US-09-298-568-1
C 8	85.2	0.9	3489	4	US-09-410-399-1
C 9	85.2	0.9	32207	2	US-08-770-379-20
C 10	85.2	0.9	32207	3	US-08-757-669A-20
C 11	85.2	0.9	32207	4	US-09-230-371A-20
C 12	81.4	0.8	124884	4	US-09-661-596A-76
C 13	79	0.8	51259	3	US-08-781-891-209
C 14	79	0.8	51259	4	US-09-618-166-209
C 15	77.6	0.8	1926	4	US-09-249-585A-2
C 16	77.6	0.8	1926	4	US-09-410-399-3
C 17	77.6	0.8	2580	3	US-09-050-863-2
C 18	77.6	0.8	2580	4	US-09-359-081-2
C 19	77.6	0.8	5452	2	US-09-130-114-1
C 20	77.6	0.8	8705	4	US-09-647-344A-14
C 21	77.6	0.8	9600	3	US-08-910-647-1
C 22	77.6	0.8	9600	4	US-09-620-925-1
C 23	77.6	0.8	10596	1	US-07-884-811-15
C 24	77.6	0.8	10596	1	US-07-885-971-15
C 25	77.6	0.8	10596	1	US-08-087-783A-15
C 26	77.6	0.8	10596	1	US-08-194-088B-15
C 27	77.6	0.8	10596	2	US-08-194-087-15

28	77.6	0.8	10596	5	PCT-US93-04648-15	Sequence 15, Appl
C 29	77	0.8	16442	3	US-08-781-891-208	Sequence 208, App
C 30	77	0.8	16442	4	US-09-618-166-208	Sequence 208, App
C 31	68.6	0.7	4549	4	US-09-620-312D-266	Sequence 266, App
C 32	68.6	0.7	4942	4	US-09-620-312D-267	Sequence 267, App
C 33	66.6	0.7	2277	1	US-08-676-967-5	Sequence 5, Appli
C 34	66.6	0.7	2277	1	US-08-676-967-5	Sequence 5, Appli
C 35	66.6	0.7	2277	2	US-09-098-487-5	Sequence 5, Appli
C 36	66.6	0.7	2481	4	US-09-894-998A-35	Sequence 35, Appl
C 37	65.8	0.7	2301	1	US-08-306-691B-23	Sequence 23, Appl
C 38	65.8	0.7	2301	4	US-09-167-206-3	Sequence 3, Appli
C 39	65.8	0.7	2301	5	PCT-US93-06251-78	Sequence 78, Appli
C 40	65.6	0.7	2721	6	5215881-2	Patent No. 5215881
C 41	65.6	0.7	8438	1	US-07-945-283-1	Sequence 1, Appli
C 42	61	0.6	150	2	US-07-829-461A-8	Sequence 8, Appli
C 43	60.6	0.6	5340	4	US-09-627-122-21	Sequence 21, Appl
C 44	60.6	0.6	15378	3	US-08-785-420-1	Sequence 1, Appli
C 45	60.4	0.6	436	4	US-09-397-787-257	Sequence 257, App

ALIGNMENTS

RESULT 1
US-09-257-179-21
; Sequence 21, Application US/09257179
; Patent No. 6410709
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 29 Human Secreted Proteins
; FILE REFERENCE: P2015P1
; CURRENT APPLICATION NUMBER: US/09/257,179
; CURRENT FILING DATE: 1999-02-25
; EARLIER APPLICATION NUMBER: PCT/US98/17709
; EARLIER FILING DATE: 1998-08-27
; EARLIER APPLICATION NUMBER: 60/056,270
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,271
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,247
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,073
; EARLIER FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 1032
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-257-179-21

Query Match 8.3%; Score 814.6; DB 4; Length 1032;
Best Local Similarity 98.3%; Pred. No. 1.2e-167;
Matches 823; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 8444 TTACTGTATCTGTAACGCCCTTATGATGAATCTAAATTTTATTTGGCTGTGATCGGT 8503
DB 179 TGTACTGTATCTGCAGACACACCTTATGATGAGTCACAAATTTTATTTGGCTGTGATCGGT 238
QY 8504 GTCAAGATTGGTACCATGGCGCTGGTGGCGATCTTGCAGAGTGGAGGAGAGCTCATTTG 8563
DB 239 GTCAAGATTGGTACCATGGCGCTGGTGGCGATCTTGCAGAGTGGAGGAGAGCTCATTTG 298
QY 8564 ATGAGTATGTCTCTCCACAGTGCAGTCAACAGAGGATGCATGACAGTCTCTCAGCCAC 8623
DB 299 ATGAGTATGTCTCTCCACAGTGCAGTCAACAGAGGATGCATGACAGTCTCTCAGCCAC 358
QY 8624 TRACAGAGAGGATTATGAGGGGTGGAAGGGGTGCTCCGTTCTTACAGGCCCATAGA 8683
DB 359 TRACAGAGAGGATTATGAGGGGTGGAAGGGGTGCTCCGTTCTTACAGGCCCATAGA 418
QY 8684 TGGCCTGGCGCTTCTTGAACACAGTACACCTTAATGATGCACAGATTATTATGGTGTTA 8743

Db 419 TGGCTGGCTTTCCTTGAACCACTAGACCTTAATGATGCACACAGATATTATTATGTTGTTA 478
QY 8744 TTAAGGAACCTATGGACCTTGCACCATGGAAGAAAGTACAAAGACGATATTATGAAA 8803
Db 479 TTAAGGAACCTATGGACCTTGCACCATGGAAGAAAGTACAAAGACGATATTATGAAA 538
QY 8804 AGCTGACGGAATTTGGCAGATATGACCAAAATTTTGTATGATCTGCTTACTACAATC 8863
Db 539 AGCTGACGGAATTTGGCAGATATGACCAAAATTTTGTATGATCTGCTTACTACAATC 598
QY 8864 CAAGTACTCCCATTTTACCAGTGTGAGAAAGTCTCGAATCATTTTGTGACAAAT 8923
Db 599 CAAGTACTCCCATTTTACCAGTGTGAGAAAGTCTCGAATCATTTTGTGACAAAT 658
QY 8924 TGAAGGCTTCAAGCTAGCAGTCTCATAAACAAACTGCACTGATACAGCTTCTTAA 8983
Db 659 TGAAGGCTTCAAGCTAGCAGTCTCATAAACAAACTGCACTGATACAGCTTCTTAA 718
QY 8984 GTTCAGCGTGTAACTTAACATAAAACACAGCAAGAAATCTGTTGCTGAACATTTTAA 9043
Db 719 GTTCAGCGTGTAACTTAACATAAAACACAGCAAGAAATCTGTTGCTGAACATTTTAA 778
QY 9044 ATTAAGGAGCCAGATGTTTGTAGTCAAGCTATCTGACAGACTTGACCTAAACTTCGTT 9103
Db 779 ATTAAGGAGCCAGATGTTTGTAGTCAAGCTATCTGACAGACTTGACCTAAACTTCGTT 838
QY 9104 TTTATTGCTCATACAGTCCCAATATATTTCTGGCCAAATTTTGTCCAAACGCAAGAAAA 9163
Db 839 TTTATTGCTCATACAGTCCCAATATATTTCTGGCCAAATTTTGTCCAAACGCAAGAAAA 898
QY 9164 AAGCAAGTCAAGACACCAATATCTTGTCAAGTACAGATGTTTACTATTGTCGAGA 9223
Db 899 AAGCAAGTCAAGACACCAATATCTTGTCAAGTACAGATGTTTACTATTGTCGAGA 958
QY 9224 AGCAGAAACTTGTATTGAAAAAAGAAAAAGAAAGCAAGAAAAAGATA 9280
Db 959 AGCAGAAACTTGTATTGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1015

RESULT 2

US-09-249-585A-4/c
; Sequence 4, Application US/0924585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

Query Match 1.0%; Score 101.8; DB 4; Length 1926;
Best Local Similarity 55.5%; Pred. No. 1.9e-12;
Matches 217; Conservative 0; Mismatches 172; Indels 2; Gaps 1;
QY 6 GGCTGAGTGGCGCCCAAGACCGCGTCTGAGTCCGCCAGGGGGGCGACGAGTACCGGAG 65
Db 760 GACGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGG 701
QY 66 GAAGCCG 125
Db 700 GAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACG 641

QY 126 AGGAGCGGGGGCGGCACACGGG--GGCGGGGGGGGGCGGGCCACCTGTCCCGGACC 183
Db 640 GACGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGG 581
QY 184 ACCGGGCG 243
Db 580 GACGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGG 521
QY 244 GAGGAAGAGGACGACATGCTCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 303
Db 520 GAGGAGGACGAGCGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGAC 461
QY 304 ACCAGAGTCTGAGACGACGAGGAGGATGAGATGGAAGAGGACGACGACGATGACTCCG 363
Db 460 GGGGAGGAGGACGAGGACGGGGAGGAGGACGGGGAGGACGGGGAGGACGGGGAGGAC 401
QY 364 TATCCGAGGAGATGGAAGACGACGACGACG 394
Db 400 GAGGACGAGCGGGGAGGAGGACGAGGACG 370

RESULT 3

US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5978807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D903051
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

Query Match 1.0%; Score 101.8; DB 2; Length 1931;
Best Local Similarity 55.5%; Pred. No. 1.9e-12;
Matches 217; Conservative 0; Mismatches 172; Indels 2; Gaps 1;
QY 6 GGCTGAGTGGCGCCCAAGACCGCGTCTGAGTCCGCCAGGGGGGCGACGAGTACCGGAG 65
Db 760 GACGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGG 701
QY 66 GAAGCCG 125
Db 700 GAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACG 641
QY 126 AGGAGCGGGGGCGGCACACGGG--GGCGGGGGGGCGCGCGCCACCTGTCCCGGACC 183
Db 640 GACGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGG 581
QY 184 ACCGGCG 243
Db 580 GACGGGAGGACGGGGAGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGG 521
QY 244 GAGGAAGAGGACGACATGCTCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 303
Db 520 GAGGAGGACGAGCGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGAGGAGGAGGAGG 461
QY 304 ACCAGAGTCTGAGGACGACGAGGAGGATGAGATGGAAGAGGACGACGACGATGACTCCG 363
Db 460 GGGGAGGAGGACGAGGAGCGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGACGG 401
QY 364 TATCCGAGGAGATGGAAGACGACGACGACG 394
Db 400 GAGGACGAGCGGGGAGGAGGACGAGGACG 370

QY 399 CAGTTACTGCACGGAAAGCAGCTTCAGGA 427
- - - - -
Db 362 CGACGACGCCATGGTTGGCATGGTCAGGA 390

RESULT 6

```

US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Egelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52368/JPW/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

	Query Match	0.98;	Score 85.2;	DB 2;	Length 3489;
	Best Local Similarity	58.88;	Pred. No. 1e-08;		
	Matches 147;	Conservative 0;	Mismatches 103;	Indels 0;	Gaps 0;
QY	198	GGCGGTCAACAAAGTGTGTACGATGACACAGCAGGAGCGCGTGGAGGAACAGGAGGA	257		
Db	1065	GGATGACGAGCAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGA	1124		
QY	258	CATGGTCTCCGAGGAGGAGGAGGAGGCGGCGACCGAGGAGACCCAGGATTCCTGA	317		
Db	1125	GGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	1184		
QY	318	GGACGACGAGGAGGATGAGATGGAAGGAGCAGCAGCATCTCCGATTATCGGAGGAGAT	377		
Db	1185	GGAGCAGCAGGAGGAGGAGGACGAGGAGGATGAGCATGATGAGACAATGAGGACGAGGA	1244		
QY	378	GGAGACGACGACGACGACCCAGTCTACTGCAACGAAAGCAGCTTCAGGAGCCCATAGTAC	437		

Db	1245	GGATGCGAGGAGGACAAAGAGGAGGACGAGGAGGACGGGGCGATGGAACAAC	1304
Qy	438	CTACAGCAGC	447
Db	1305	GTTGAGCATC	1314

RESULT 7

```

US-09-298-568-1
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 18412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1

```

```
Query Match          0.9%; Score 85.2; DB 4; Length 3489;
Best Local Similarity 58.8%; Pred. No. 1e-08;
Matches 147; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
```

QY	198	GGCGCTCAACAAAGTGGTTACGATGACCCACGAGCGAGCGCGTGTGAGGGAAGAGGAGGA	257
Db	1065	GGATGACGAGGAGGATGACGAGGAGGATGACCGAGGATGACGAGGAGTATGACGAGGA	1124
QY	258	CATGCTCTCCGAGGAGGAGGAGGAGGACGGCGACGCCGAGGAGAGACCCACGAGTTCTGA	317
Db	1125	GGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGACGAGGAGGAGGACGAGGA	1184
QY	318	GGACGACGAGGAGGATGAGATGGRAGAGGACGACGATGACATCCGATTATCCGGAGGAGAT	377
Db	1185	GGAGGACGAGGAGGAGGAGGACGAGGAGGATGACGATGATGAGGACATATGAGGACGAGGA	1244
QY	378	GGAACGACGACGACGACGCCAGTCTTCTGCACGGAAAGCAGCTTCACGAGGCCATGATPAC	437
Db	1245	GGATGACGAGGAGGAGGACAGAGGAGGACGAGGAGCGGGGCGGATGGAACAAAC	1304
QY	438	CTACAGCAGC	447
Db	1305	GTGAGCATC	1314

RESULT 8

```

US-09-410-399-1
; Sequence 1, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; TITLE OF INVENTION: to Genomic Host DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus

```

US-09-410-399-1

Query Match 0.9%; Score 85.2; DB 4; Length 3489;
Best Local Similarity 58.8%; Pred. No. 1e-08;
Matches 147; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 198 GGCCGTCACAAAGTGGTGTACGATGACACACGAGCGAGCGGCGGTGGAGAGAGGAGGA 257
DB 1065 GGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGA 1124
QY 258 CATGGTCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATCTGA 317
DB 1125 GGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 1184
QY 318 GGACGACGAGGAGGATGATGATGAAGAGGACGACGATGACTCCGATTATCCGAGGAGAT 377
DB 1185 GGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATGATGAGGACAATGAGGAGGAGGA 1244
QY 378 GGAGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATAGTAC 437
DB 1245 GGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 1304
QY 438 CTACAGCAGC 447
DB 1305 GTTGAGCATC 1314

RESULT 9

US-08-770-379-20/c

; Sequence 20, Application US/08770379

; Patent No. 5849564

; GENERAL INFORMATION:

; APPLICANT: Chang, Yuan

; APPLICANT: Bohenzky, Roy A.

; APPLICANT: Russo, James J.

; APPLICANT: Edelman, Isidore S.

; APPLICANT: Moore, Patrick S.

; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED

; HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Cooper & Dunham LLP

; STREET: 1185 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/770,379

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: White, John P.

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 52342

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 278-0400

; TELEFAX: (212) 391-0525

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3207 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-770-379-20

Query Match

0.9%; Score 85.2; DB 2; Length 32207;

Best Local Similarity 58.8%; Pred. No. 3.2e-08;
Matches 147; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 198 GGCCGTCACAAAGTGGTGTACGATGACACACGAGCGAGCGGCGGTGGAGAGAGGAGGA 257
DB 20932 GGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGA 20873
QY 258 CATGGTCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATCTGA 317
DB 20872 GGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 20813
QY 318 GGACGACGAGGAGGATGATGATGAAGAGGACGACGATGACTCCGATTATCCGAGGAGAT 377
DB 20812 GGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATGATGAGGACAATGAGGAGGAGGA 20753
QY 378 GGAAGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATAGTAC 437
DB 20752 GGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 20693
QY 438 CTACAGCAGC 447
DB 20692 GTTGAGCATC 20683

RESULT 10

US-08-757-669A-20/c

; Sequence 20, Application US/08757669A

; Patent No. 6183751

; GENERAL INFORMATION:

; APPLICANT: Chang, Yuan

; APPLICANT: Bohenzky, Roy A.

; APPLICANT: Russo, James J.

; APPLICANT: Edelman, Isidore S.

; APPLICANT: Moore, Patrick S.

; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Cooper & Dunham LLP

; STREET: 1185 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/757,669A

; FILING DATE:

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: White, John P.

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 45185-F

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 278-0400

; TELEFAX: (212) 391-0525

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3207 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-757-669A-20

Query Match

0.9%; Score 85.2; DB 3; Length 32207;

Best Local Similarity 58.8%; Pred. No. 3.2e-08;
Matches 147; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Db	783	AGGAGCAGGAGGGCCAGGAGGGCCAGGAGGAGGGGCGAGGAGGGGCGAGGAGGAGG 842
Qy	218	ACGATGACCAAGAGAGCGCGGTGGAGGAAGAGGAGGACATGGTCTCCGAGGAGGAGG 277
Db	843	AGGAGGGCCAGGAGGGCCAGGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGGGCGAGG 902
Qy	278	AGGAGGAGGAGGGCGCGCCGAGGAGACCCAGGATTCTGAGGACGACGAGGAGGATGAGA 337
Db	903	AGGGCCAGGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGG 962
Qy	338	TGGAAGAGGAGGAGGATGACTCCGATTATCCGAGGAGAGATGGAGAGGAGGAGGAGCG 397
Db	963	GGCAGGAGCAGGAGGTGGAGCGGGGTCGAGGAGGAGCTGGAGGCGGGGTCGAGGAGG 1022

Search completed: September 23, 2003, 09:19:26
Job time : 408 secs

Matches 339; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 7531 CTGCAACAACAAGTCCAGGTTCTCTCAGATCCAGTCCAGGTTGTGGCTCAGATACAG 7590

Db 341 CTGCAACAACAAGTCCAGGTTCTCTCAGATCCAGTCCAGGTTGTGGCTCAGATACAG 282

QY 7591 GCTCAGCAAGTGGTGTGCCCCAGCAAAATCAAACTCCAGTTACCTATCCAAATTCAGCAA 7650

Db 281 GCTCAGCAAGTGGTGTGCCCCAGCAAAATCAAACTCCAGTTACCTATCCAAATTCAGCAA 222

QY 7651 AGCAGTGTGTGCAGACTCACCAGATTCCAGATGTCAGATGTCAGGTCAGGCGCCAGTGG 7710

Db 221 AGCAGTGTGTGCAGACTCACCAGATTCCAGATGTCAGATGTCAGGTCAGGCGCCAGTGG 162

QY 7711 CAAGAGCAGTGTGCAAGGGTTCAGCAACTCAGGATCAGCAGCAAAAGAAAGCAAGCA 7770

Db 161 CAAGAGCAGTGTGCAAGGGTTCAGCAACTCAGGATCAGCAGCAAAAGAAAGCAAGCA 102

QY 7771 CAGATAGAAATTA-ACGTGAACACACACCCCTCCAGCTTCTAATCAAAAGTTGAATCATTC 7829

Db 101 CAGATAGAAATTAAGCGGTGAACACACCCCTCCAGCTTCTAATCAAAAG-TGAATCATTC 43

QY 7830 GAAACAGGTGGTGATGAGCATAAATGCTGTAATAGAACATTT 7871

Db 42 GAAACAGGTGGTGATGAGCATAAATGCTGTAATAGAACATTT 1

RESULT 13

US-09-796-692-8224

; Sequence 8224, Application US/09796692

; Publication No. US20020198362A1

; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander

; APPLICANT: Mannion, Jane

; APPLICANT: Algate, Paul A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

; FILE REFERENCE: 2077.001200

; CURRENT APPLICATION NUMBER: US/09/796,692

; CURRENT FILING DATE: 2001-03-01

; PRIOR APPLICATION NUMBER: 60/186,126

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: 60/190,479

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: 60/200,545

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: 60/200,303

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/200,779

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/200,999

; PRIOR FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 60/202,084

; PRIOR FILING DATE: 2000-05-04

; PRIOR APPLICATION NUMBER: 60/206,201

; PRIOR FILING DATE: 2000-05-22

; PRIOR APPLICATION NUMBER: 60/218,950

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: 60/222,903

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: 60/223,416

; PRIOR FILING DATE: 2000-08-04

; PRIOR APPLICATION NUMBER: 60/223,378

; PRIOR FILING DATE: 2000-08-07

; NUMBER OF SEQ ID NOS: 9597

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 8224

LENGTH: 545

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: unsure

LOCATION: (419)

; OTHER INFORMATION: n=A,T,C or G

US-09-796-692-8224

Query Match 2.3%; Score 223.2; DB 10; Length 545;

Best Local Similarity 96.2%; Pred. No. 1.4e-40;

Matches 228; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 8444 TTACTGTATCTGTAAGCGCTTATGATGAATCTAAATTTTATATTTGCTGTGATCGGT 8503

Db 309 TGTACTGTATCTGCAACACACCTTATGATGAGTCACAATTTTATATTTGCTGTGATCGGT 368

QY 8504 GTCAGAAATTTGATACATGGCGCTGCGTTGGGATCTTTCAAAAGTGAGGCGAGCTCATTTG 8563

Db 369 GTCAGAAATTTGATACATGGCGCTGCGTTGGGATCTTTCAAAAGTGAGGCGAGCTCATTTG 428

QY 8564 ATGAGTATCTGTCACAGTCCAGTCAACAGAGATGCCATGACAGTGTTCACGCCAC 8623

Db 429 ATGAGTATCTGTCACAGTCCAGTCAACAGAGATGCCATGACAGTGTTCACGCCAC 488

QY 8624 TAACAGAGAAGGATTATGAGGGGTTGAAGAGGTTGCTCGTTTCCTTACAGGCCCAT 8680

Db 489 TAACAGAGAAGGATTATGAGGGGTTGAAGAGGTTGCTCGTTTCCTTACAGGCCCAT 545

RESULT 14

US-10-040-862-8224

; Sequence 8224, Application US/10040862

; Publication No. US20030078396A1

; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander

; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane

; APPLICANT: Retter, Marc

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther

; FILE REFERENCE: 014058-01352005

; CURRENT APPLICATION NUMBER: US/10/040,862

; CURRENT FILING DATE: 2001-11-06

; PRIOR APPLICATION NUMBER: US 60/186,126

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: US 60/190,479

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: US 60/200,545

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: US 60/200,303

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: US 60/200,779

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: US 60/200,999

; PRIOR FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: US 60/202,084

; PRIOR FILING DATE: 2000-05-04

; PRIOR APPLICATION NUMBER: US 60/206,201

; PRIOR FILING DATE: 2000-05-22

; PRIOR APPLICATION NUMBER: US 60/218,950

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: US 60/222,903

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: US 60/223,416

; PRIOR FILING DATE: 2000-08-04

; PRIOR APPLICATION NUMBER: US 60/223,378

; PRIOR FILING DATE: 2000-08-07

; PRIOR APPLICATION NUMBER: US 09/796,692

; NUMBER OF SEQ ID NOS: 10467

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 8224

LENGTH: 545

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: unsure

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 01:09:02 ; Search time 22.9984 seconds
(without alignments)
5348.092 Million cell updates/sec

Title: US-09-698-295-1
Perfect score: 14971
Sequence: 1 MWSEEEEDGDAETQDSE.....KLGFKASRSHNNKLOSTAS 2907

Scoring table: BLOSUM62
Gapop 10.0 ; Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	982.5	6.6	238	4	US-09-257-179-80
2	491	3.3	2137	4	US-09-134-001C-4463
3	441.5	2.9	3636	4	US-09-134-001C-5080
4	391	2.6	10182	4	US-09-134-001C-3159
5	386.5	2.6	2136	4	US-09-854-856-14
6	380.5	2.5	1999	4	US-09-854-856-16
7	380.5	2.5	2108	4	US-09-854-856-30
8	380	2.5	3878	4	US-09-914-259-11
9	378	2.5	8991	4	US-08-714-741-32
10	376	2.5	2076	4	US-09-854-856-46
11	374.5	2.5	1971	4	US-09-854-856-32
12	374	2.5	2201	4	US-09-854-856-24
13	370	2.5	1939	4	US-09-854-856-48
14	370	2.5	2048	4	US-09-854-856-62
15	369	2.5	2229	4	US-09-854-856-8
16	368	2.5	2064	4	US-09-854-856-26
17	367	2.5	3969	3	US-08-061-376-5
18	364	2.4	1911	4	US-09-854-856-64
19	363.5	2.4	2141	4	US-09-854-856-56
20	363	2.4	2092	4	US-09-854-856-10
21	362.5	2.4	2842	1	US-07-741-940-7
22	362.5	2.4	2842	1	US-08-289-548A-7
23	362.5	2.4	2842	1	US-08-452-654-7
24	362.5	2.4	2842	1	US-08-449-731-7
25	362.5	2.4	2843	1	US-08-452-655B-2
26	362.5	2.4	2843	1	US-08-452-655B-7
27	362.5	2.4	2843	3	US-08-450-582-2

28	362.5	2.4	2843	3	US-08-450-582-7	Sequence 7, Appli
29	362.5	2.4	2973	2	US-08-821-355A-7	Sequence 7, Appli
30	362.5	2.4	2973	2	US-09-003-687A-7	Sequence 7, Appli
31	362.5	2.4	2973	3	US-09-136-605-7	Sequence 7, Appli
32	361	2.4	2843	1	US-07-741-940-2	Sequence 2, Appli
33	361	2.4	2843	1	US-08-289-548A-2	Sequence 2, Appli
34	361	2.4	2843	1	US-08-452-654-2	Sequence 2, Appli
35	361	2.4	2843	2	US-08-370-235A-2	Sequence 2, Appli
36	361	2.4	2843	4	US-08-449-731-2	Sequence 2, Appli
37	358.5	2.4	2169	4	US-09-854-856-40	Sequence 40, Appli
38	357.5	2.4	2004	4	US-09-854-856-58	Sequence 58, Appli
39	354.5	2.4	3248	1	US-08-353-700-1	Sequence 1, Appli
40	354.5	2.4	3248	5	PCT-US93-16216-1	Sequence 1, Appli
41	352.5	2.4	2032	4	US-09-854-856-42	Sequence 42, Appli
42	352.5	2.4	2035	1	US-08-046-585-5	Sequence 5, Appli
43	352.5	2.4	2035	1	US-08-393-703-5	Sequence 5, Appli
44	352.5	2.4	2035	5	PCT-US93-11721-5	Sequence 5, Appli
45	345.5	2.3	2354	4	US-09-854-856-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1
US-09-257-179-80
; Sequence 80, Application US/09257179
; Patent No. 6410709
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 29 Human Secreted Proteins
; FILE REFERENCE: P2015P1
; CURRENT APPLICATION NUMBER: US/09/257.179
; CURRENT FILING DATE: 1999-02-25
; EARLIER APPLICATION NUMBER: PCT/US98/17709
; EARLIER FILING DATE: 1998-08-27
; EARLIER APPLICATION NUMBER: 60/056,270
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,271
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,247
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,073
; EARLIER FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-257-179-80

Query Match 6.6% Score 982.5; DB 4; Length 238;
Best Local Similarity 78.8%; Pred. No. 1.7e-51;
Matches 189; Conservative 7; Mismatches 19; Indels 25; Gaps 3;
Qy 2689 PTLVASQKRREREK-----DSSSKSKKKMMISTTKTKK-----DT 2727
Db 3 PTPR-----KTPYDESKFYIGCDLCTNWHGECVITEKEAKMDVYICNDCKRAQEGSSE 58
Qy 2728 KLYCICKTPYDESKFYIGCDRCQNWYHGRGVIGLQSEAEILIDEVYVCPQOCSTEDAMTVLT 2787
Db 59 ELYCICKTPYDESKFYIGCDRCQNWYHGRGVIGLQSEAEILIDEVYVCPQOCSTEDAMTVLT 118
Qy 2788 PLTEKDEGLKRVLSLQAHKMAWPFLEVPDPNDAPDYGVYIKFPMDLATMEERVQRRYY 2847
Db 119 PLTEKDEGLKRVLSLQAHKMAWPFLEVPDPNDAPDYGVYIKFPMDLATMEERVQRRYY 178
Qy 2848 EKLTFFVADMTKTFDNCRYNPSDPFYQCAEVLESFFVOKLGFKASRSHNNKLOSTAS 2907
Db 179 EKLTFFVADMTKTFDNCRYNPSDPFYQCAEVLESFFVOKLGFKASRSHNNKLOSTAS 238
RESULT 2

US-09-134-001C-4463
; Sequence 4463, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 3.3%; Score 491; DB 4; Length 2137;
Best Local Similarity 17.7%; Pred. No. 1.3e-20;
Matches 410; Conservative 357; Mismatches 1006; Indels 546; Gaps 74;

QY 345 ENENEKKIWIYSTKQVLAELIDCLDKDYWEAEL--CKILEEMREIEHHRMDITEDLTNKA 402
DB 12 EDYKASLEQQKTRVKIYK---SGKSVKASINIEILLKTMGLPFLSKNEIQENVTEXT 67
QY 403 RG-----SNKSFLAAANEIELESIRAKKGDIDNVKSPETEK 439
DB 68 KGHKLKSAKTTALVGGAFTFNMNLNNHQAFASETPTITSEISSNSETVANQSTTIKNS 127
QY 440 DNEIENDESKDAEKNREEDFOSLEKSDDDKTPDDPEQCKSEEBTEVCDKGNV---S 495
DB 128 QKETVNSTLESHNSNSTKQMSSEVTNDQSE---KAGISQQSSETSNOSSELTAYAS 184
QY 496 ANLGDMTNTATSEETSPSGRSPVGCISPTDSSNNAKKVASELPQDV-----P 545
DB 185 TDHVESTTNNDAQAQDNKSNVTSKTSQNTSSSEKNISSLTQSTETKATDSLATS 244
QY 546 EEPNKTCESSNTSATTSTQ-----PNLE-----NSNSSELNNSQSBSA 585
DB 245 EARTSTNOISLNTSTSTNSQSSPTSPANLRTSFRTVLNTMAAPTSTTTTSSLTNSV 304
QY 586 KAADDPENGERESHTPVSIQEEIVGDFTEKSTGELSESPGAGKASGSTRITILRNPD 645
DB 305 VNKDNFN-----EHNLS-----GSATYDPKTGATLPDA-YSOKGAISLTRL--- 349
QY 646 SKLSQLKSOQVAAAAHEANKLPKEGKVLVN-----SQGEISRL 685
DB 350 -----DSNRSFRFTGKVLGNRYEGYSPDGTGTDGIGAFSPGLGQI 393
QY 686 STKKEVIMKGNINNYFLQEGKYRYHHQYNTSNFALNKHQRE-----DH 732
DB 394 GKECAAVGIGLGNAFGF---KLDYHNTSTPKDAKADPRNVGGGAGAFGVSTDR 449
QY 733 D-----KRRHLAHKFLTPAGE-----FKWNSVHGSKVLITSLTLITQLENNIPSSF 782
DB 450 NGMATTEASAAKLVNQPTDNSQDFVIDYNGD---TKVMTYIAGQTETR----- 497
QY 783 FHPNASHRANWTIKAVOMCSKPREFALALAECAVKPVVMLPIWREFLGHTRLHRMTSI 842
DB 498 -----NLTDWIKN-----SGGTTSLSMSTAGGAKNLQOVQFGTFEYTESAVAKVRVY 546
QY 843 EREE-KEKVKKKKEKEEETMQOATWXYTPPVKHQV--WKQGEYRVT-----GYG 893
DB 547 DANTGKDITPPKTIAGEVDATV-----IDKQLNKLNSGYSVSTDALQNSYTS 596
QY 894 GWSWI-----SKTHVRFVFKLPNGNTNVNRYKSLGKTKNMMDENMDESCKKCRSP 945
DB 597 ETSCTPLKLTNSSQIVYKFN-----KDVQGPQISVD-----SQTREVGKTI 638

QY 946 KKIKI-EPDSEKD-----EVKGSDAAGADQ-----NEMDISKITEKKDQOVKELL 991
DB 639 NPITITTTDNSKQDLVTTTGLPSGLSFDQTTNTITCTPSEVGGTTTIT-----VNTTD 691
QY 992 SDSKPKCKEPMEDDDMKTESHVNCQESSQV-----DVVNVSE 1030
DB 692 ATGNVTSKQFTITITQDTPSPVNVTPQASEVFPINPITITATDNGKVVVHTVTGLPQ 751
QY 1031 GFHLRTSYKKKTKSKLDGLLEKRIKQFTLEEKORLEKIKLEGKIGKGTSTNSSKNLS 1090
DB 752 GLKFDASTNSIVGTPQIGINTIETDASGNKTTTKINYE-----VTRNSASDST 803
QY 1091 ESPVITKAKGECQ-----SDSMRQEQSPNANNQDPELQCCGSDSSVLRMSDPSTHTN 1145
DB 804 STSIVNSVSTISNSSTLSDSVKASQSLSTKSLSLSESL--SASTSNSTQASEASTSK 861
QY 1146 KLYPKDRVLDDVYSIRSPET-----KCPKQNSIENDIEEKVSDLA---SRQEPKTKSK 1196
DB 862 QL--SESASTSTSDSASEARKSESTSKTSLSLSESTSTSVSDSASVSTSESASTSTSV 919
QY 1197 GNDFFIDDSKLASADDIGTLICKNNKPLIQESPTIVSSKSLALHSSVPKSTNDRDATPL 1256
DB 920 STTSISIDSTSTSTSDSASIKASEASTSKLSESVSTSTSDSASTSTSVSDSNASTSL 979
QY 1257 SRA-----MDPEGLGCDSESNSTLENSDTSVSIQDSEEDMIVQNSIESIQFRTREQ 1311
DB 980 SKSTSTSVSDSTSTSTSDSASTSTSESDSASTS-----LSESTSTSVSDSTSTSTS 1032
QY 1312 DVEVLEPLKCE-----LVYSGESTGNCEDRLPVKGTGANGKPKSQOKLEER 1357
DB 1033 DSASMSASESBNKSKTSLSLSESTSTLSGTSASTSDSASTSTSES--ESDSTSTSLSES 1090
QY 1358 PVNKCSDQIKLNTTKNNENRESEKKGQRT-----STQFIN 1395
DB 1091 TSTSLSGTS--ASTSDSASTSTSESDSTSESTSLSESLSTSVSDSTSASTSEASTSTSE 1149
QY 1396 GKDNPKPIYLLGECLEKSESRVGVNPEKPVNNINKIIPENDIKSLTVKESAIRPEFNG 1455
DB 1150 SESNASTSLSGSLSTSDSTSTSDSASTSTSES---ESDSTSTSLSEST----- 1199
QY 1456 DVIMEDFERNSSSTFKSHLLSSDAEGNRYRDSLETLPSTKESDSTQTTTPSACPEGNSV 1515
DB 1200 STLSLSDSTSTSESASTSTSESDDSTSE-STLSLSESTSTSVSDSTSAST--SDSASTSTSV 1257
QY 1516 NOVEDMETETSE-----VKVYTS-----SPITSEES-----NLNDFIDENGL 1554
DB 1258 SDSESASTSISELSTSVSDSTSTSDSASTSTSESDSTSESTSLSESLSTSVSDSTSA 1317
QY 1555 PINKNENVN-GESKRKTVITVTTMTSTVATESKTIVKEKGDQTVVVSSTENCAKSTVT 1613
DB 1318 STSDSASTSTSESDSASTSLSGSTSTSLSDSTSTSDSASTSTSESDSERASTSLSG 1377
QY 1614 TTTTTVTKLSTPSTGGSVDDIISVKBQSKTVVTVTTVDS---LTTTGGTLVTSMTVSKYS 1670
DB 1378 STTSLSLSDSTSTSTSDSAS-----TSTSVSDSNASTSLSGSLSTSVSDSTSTS 1426
QY 1671 TRDKVKLMKFRPKKTRSGTALPSYRKFTVTKTKSIFVLNDDLLKLARKGGIREVPYF 1730
DB 1427 TSDSAS--ASTSESDSERASTSL-----SGSTSTSI-----SDSTSTSTSDSASTSTSV 1474
QY 1731 NYNAKPALDIWPPSPRPTFGITWRYLQTVKSLAGVSLMLKLLWASLRWDDMAKPPG 1790
DB 1475 ESN-----TSTSI-----ESLSTSV--- 1491
QY 1791 GGSTRFETSETITTEIKRDRVGPYGRFPYCIKIKICPIGVDETPEKPTPORKGLR 1850
DB 1492 SDSTSTSDSASTSTSVSDSASTSSSE-----SVSTSDSESTSTSDSAST 1541
QY 1851 SSALRPKRPETPKQCPVLIETWVAEELELWEIRAFARVEKEKAQAEQAKKRLBOQ 1910
DB 1542 STSVSESNSTSLSGST--STSVSDSTSTSTSDSASTSESD-----SDSASTSSSES 1594
QY 1911 KPTVIATSTSTSTSTSTISPAQKVMVAPISGSVTTGTGMVLTATTKVGSPTATVTFQKNK 1970

Db 1595 VTSVSDSTGASTSESASTSTSVSDSNASTSLSESTSTSLSDSTSMST----- 1643
Qy 1971 FHOTFATVWQOGSGVVOQKVLGIIPSSTSQOTFTSFOPRTAT-----VTIR 2023
Db 1644 -SDSASTSTSESDDASTSLSDSTSTSVSESTSTSTSVSASNSTSTSLSDSRSTSL 1702
Qy 2024 PNTSGS---GTTSTNSOVIITGPQIRPGMTVIRPLOQSTLGKALIRTPVMVQGAPOQVM 2080
Db 1703 DSTSTSTSESGSTSE-----SDSDSASTSLSEST-----S 1734
Qy 2081 TOIRGQPVSTAVSAPNTVS---STPGQKSLTATSTVS-----NIOSSASOPPRPOQOV 2132
Db 1735 TSTSDSTSTSDASTSMVSUSNRASTSLSDSTSTSVSDSTSTASTSESASTSTRES 1794
Qy 2133 KLTMAQTLTQGHGNGQTLVIQGGQTTGQLLIPQGVTVLPFGPQOOLMOAAMPNGT 2192
Db 1795 TSASTSLSESTSTSVSDSTSTSDASTSTSESSTSESDSTSTSVSDSTSTSVSDST 1850
Qy 2193 VQRLFTPLATTATTATTTTSTVTTAAGTGQRQSKLSPQMQLPQVQDQKTLPPAQSSVGP 2252
Db 1851 -----ASTSASTSTSVSDSNAST-----SLSGSTSTSVSD 1883
Qy 2253 AKAPOTAPQASRPPQTPQSPAQPEVQTPQEVQTTQTVSSHVPSAQTHAQSSKPOV 2312
Db 1884 STSTSTASASTSTSESDDASTSLSGSTSTSLSDSTSTSDASTSTSESASTSTSV 1943
Qy 2313 AAOQPQSNVQOGSPVQVQSPQSOT-----RIRPSTPSQLSPGQSQVOTTTSQPIQPHT 2368
Db 1944 SEDSESTSVSESTSTSVSDSTSTSTSESASTSTSESESTSESTSVSESTSTSDSST 2003
Qy 2369 SLQIPSGQSPQVQVQSSQTQLSSGQTLNOVSVSPSR 2407
Db 2004 STSMSTSTETSPINSSEQFI--GDSLSEDTIVTQSK 2040

RESULT 3

US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08 60/064, 964
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 2.9% Score 441.5; DB 4; Length 3696;
Best Local Similarity 17.1% Pred No. 2.7e-17;
Matches 471; Conservative 470; Mismatches 1066; Indels 749; Gaps 106;

Qy 311 QKNKPYIRHPIGYD-RSRKKYWLNR--LIIEEDTENENEKKIWIYSTKVQLAELIDC 367
Db 569 QVNKDII---PSNYTLASYNKYNKLERAQTVLDEETNTP----- 606
Qy 368 LKDYWEAEUCKILEMREIEHHRMDITDLTNKAGNSKSFLLAANEETLESIRAKKGD 427
Db 607 FNORYSQTQIDLLHELOTTPLINRVSASREINDKA-----QEMTDV----- 648
Qy 428 IDNVKSPETEKKNETENDSKAEKNREEFEQSLSEKSDDKTPDDDPQGGK----- 480
Db 649 -----YDSTELTTEKOTLVQIENHRNEISNNIDDELTDGVERVKEAGLHTL 697

Qy 481 -SEEPTEV-----GDKGNSVSNALGDNTTNTATSE-----ETSPSEGRSPVGC 522
Db 698 ESDTPHPVTKPNARQVNNRADQOKTLIRNHNHEATTEQNEAIRQVEAHSSDAIAKIG-E 756
Qy 523 SEPPDSNMAEKKVASLPEQDVPPEPKTCESSTNTSATTTSIQNLNLSNSSE----- 576
Db 757 AETDTTNEARDNGTKLIATDV---PNPTKAEARAAVTNSANSKIDINNTOATLDER 813
Qy 577 -----LNSOSE-----SAKAADD-----PENG-----ERESHPTPVSIQEEIV----- 609
Db 814 NDATALVNRKDEAIQINNTAQGNDDVTEAGNNGTNTIQOVPLTPVPRQNAIATINAKAD 873
Qy 610 -----GDTFSEKSTGELSSEPGAGKASGSTRITRLRNPDSKLSQLKSOQVAAA 659
Db 874 EQRLIQANNNATTEKADAERKVNEAV---ITANQNITNATTNRDVOAQTTGSGIISA 930
Qy 660 AHEANKLFKECK-----EVLVVSNOGELISRLSTKKEVIMKGNINNYFKLOGEGKYRVHN 714
Db 931 ISPATKIKEDARAAVEAKATAQNOQINSNNMATTEE---KDALNQVEAHKQAAIATINQ 987
Qy 715 QYSTNSFALNK-----HQHREDHKRRHLAHFCLTPAGEFKWNGSVHGSKVLTISTL 767
Db 988 AQSTQVSEAKNNGINTINQDOPNAVKNNT--KTILEQKGNEK-KSAIAGTPDATTEEK 1044
Qy 768 RLITITOLENNIPSSFFPHNMAHRANWIKAVQMCSPREFALALAILCAVAPVYMLPIW 827
Db 1045 QEAVSAVSQAVTNGITHTINQANSND-----VDQELSNABEQIITHTNVNVQKKPQA 1095
Qy 828 REFLGHTRLHRWTSIEREEKEKVKKKEK-----KQEBEETMQQATWYKTFPVKHQV 879
Db 1096 QALIAKTNEKQSAINDNEGTIEEKQAIQSLNDAKLADEQITQAA----- 1143
Qy 880 WKQKGEYRVTVGYGGSWISKTHVYRFVPLPGTNTVNYRKSLSGTGNNMDENDESKR 939
Db 1144 ---SNQVNDNALNIGISNISKIQT-NFTKQQAQDQVNOQFQEAELNSTPHATQDEKQ 1199
Qy 940 KC---SRSPKKIKIEPDSEKDEVKGSDAAKGADQNMEDIKITEKQDQVDELLSDSDK 996
Db 1200 DALTRLTOAKETALNDINQACTNQVDTALTSGIONIQNTQVNVKQKQEAKTII-NDIVQ 1258
Qy 997 PCKEEPEVDDDKTESHV--NCOESSQVDVNV-----VSEG-----FHL 1034
Db 1259 QHKQSIQNDQDATTKEEVEANVLNVAQNVISKIDNATTNNQIDGIVSGRQINATTP 1318
Qy 1035 RTSVK-----KTKSSKLDGLLE-----RRIKQFTLEKORLEKIKLEGG 1074
Db 1319 DTSIKRNAKNDIDIKAAADKKIKIQINDATDEETQEBANRKEIEAKIENIQNRSTRDQ 1378
Qy 1075 IKGIGKTSTNSSKNLSESPVI-TRAKEGCSDSMRQ---EOSPNNANDQPEDLIQCSQ 1129
Db 1379 VNEAKTNGINKIENITPATTVKSEARQAVQNKANEQINHQTQNTPDATNEEKQEAIRVSA 1438
Qy 1130 SSSVLRMSDPSTHTNKLKPDRLVLD--VVSIRSPETKCPKQNSIENDIEKVS----- 1181
Db 1439 ELARVQAQINAEHTTQGV---KTIKDDAITSLSRINAQVYVEKESARNAIEQKATQQTQFI 1495
Qy 1182 -----DLASRGQEPKSK-----TKGNDFDFFDDSKLASADDITGLI----- 1217
Db 1496 NNDNATDEKEVANNLVIAATKQSLDNINSLSNND--VENAKVAGINEIANYLPATAV 1553
Qy 1218 -CKNKPLIQEESDTIVSSKSAHSSVPKSTNDRDAT-PLSRAMDFEGKLGCDSE--SN 1273
Db 1554 KSAKAKDIDQKLAQI-----NQIOTHOTATTEKEAAIQLANQKSNEARTAIQNEHSNN 1608
Qy 1274 STLSSSDTYS-----IODSSEEDMIVQNSNESISEQFR-----TRQDVEVLPLK 1320
Db 1609 GVAQAKSNGIHEIELVMPDAHKKSDAKQSDKNRYNEQSNINTTPDPTDEEKQKALDKL 1668
Qy 1321 CELVSG-----ESTGNCEDRLPVKGTGAN-GKKPSQOKKLEER----- 1357
Db 1669 IAKDAGYNKYDQACTNOQVSDAKTEALDITNTIQANVAKRPSARVELDSKDFEDLKROINA 1728

```
QY 1358 -----PVMKCSQIOLKNTTDDKNNENRESEKK-QRTSTFOI----- 1394
D 1729 TPNATEEKBQDAIORLNGRDEVKNLINOQRRDNEVEQHKNGIGOLELTHANPTRKSDA 1788
QY 1395 -----NCKDNKPKIYLGECLEKESSESVYGNVPEKVN----- 1429
D 1789 LOELQTKFISQTELINNKDADNEEKDEAKRLLEISKNTTINNOAGTNNOVDNAKDN 1848
QY 1430 ---INKIIPENDIK-----SLTVKESAIRPFING--DVIMED-----FNERNSEKSH 1473
D 1849 MNEIATILPATIKTDAIDKAEQOQVTIINGNDNATDEKAEARKLVEKAKIEAKSN 1908
QY 1474 LLSSS---DAENGYRDSLETLPSTKESDSTQTPSPASCPESNSVNOVEDMEIETSEVK 1530
D 1909 ITNSDTEREVNGAKTNGLEKINNIQPSOTKIT-----NAKQELNDAKAE--QLIQ 1956
QY 1531 VTSSPITSSEENSLNDFIDENGLPINKNENNGESKRKTIVITEVTMTSTVATESKTVI 1590
D 1957 INNTPDTEEBKQEAATNRVNA-GL-AQAIONINNAHSTQEVNESKTSIATIKSVOPNVI 2014
QY 1591 K-----VEKGDKQTVVSTEN-----CAKSTVT-TTITVTKLSTPSTGSGVDI 1633
D 2015 KKPTAINSLTQEAANNQKTLIGNDNATDEKAEAKQLVTKLNEIOIQIHSTQDNQVD- 2073
QY 1634 ISVKEQSKVTWTTVTDS-----LTTTGGTILVTSMVTSKEYSTRDKVKLMKFSRPK 1684
D 2074 -NVKAGQITAIKILINAHAKRQDAINILTNLAESKSDIRANODATTEK----- 2122
QY 1685 KTRSGTALPSYRKFTVKSTKSFIVLPNDLKKLARKGIREVPYFNYNNAKPAIDWYP 1744
D 2123 ---NTAIQSI-----DQTLAQAR-----NNGINGANTNA----- 2147
QY 1745 SPRPTGTYWRYPLQIVKSLAGVSLMLRLWASLRWDMMAKVPVGGGSTRTSETIT 1804
D 2148 -----LVDENLEDGQKQRLVILSTQTKQAKADIAQAIGQORSTIQNQNA 2194
QY 1805 TT---EIKR-----RDVPGYI---RFEYCI 1825
D 2195 TTEKQEAERLNAQETNGVNDRIOAALANQNTDEKNILETIRNVEPIVIVKPRANEII 2254
QY 1926 RKIICPGVETPKETPTQKGLRSALRPRKPTKQTPGVIIETWVAEELELEWEIR 1885
D 2255 RKAAEQTTILNQDQATLEEKQIALGLEEVKEALNQ-----VSAHNSNDVKIAENN 2309
QY 1886 AFARVEKKAQAQAEQAKRLQKQPTVIATSTTSPTSTSTISPAQKVMVAPIS--G 1943
D 2310 GIAKISEVHPETIKNKAQEIQDAQSQIDITNANKSTNEEKSAADRVNVAKIDAIN 2369
QY 1944 SVTTGKMWLTKVGSFATVTFQONKNFHOTFATVWKQGSNGVGVQVQKVLGIIPSS- 2002
D 2370 NITNAT---TTQLVNDA-----KNSGNTSISQ-----ILPSTA 2399
QY 2003 --TGSQOTFTSFQPRATVTIRPNTSGSGTTSNOVIRPGPQIRPGMTVIRPPLQO--- 2057
D 2400 VKTNALAALASEAKKNKAIIDQTPNATAEKEEANKV-----DRLOEAD 2445
QY 2058 STLKAIIRTPV-MVOPGAPQQ---VMTQIIRQGPVSTAVSAPNTVSSTPGQSKLSATS 2113
D 2446 ANILKAHTTDEVNINKQAQVQINAVQVEIKQNVKNQLN-----QFIDNQKKIIENTP 2500
QY 2114 TSNTOSSASQPPRQPGQGVKLTMAQLTQCGHGGNGLVIVTQGGQTTGQLQLIPGV 2173
D 2501 DATLEKA-EANRLLOLVLTSTSDSIANDVHNNEVDQALD-----KARKPIE 2546
QY 2174 TVLPCGQOLMQAAMPNGTVQRFLETPPLATTATTSTTTTSTVSTAAGTGEQKQSKLSPQ 2233
D 2547 EIVP---QVSKRRDLVNAIQEAF-----NSQTQEIQENQEAATNEEKTEALNKI 2591
QY 2234 MQ-VHQDK-TLPPAQSS--SVGPAKAPQTAOPSARPOPTQOPSPAPQEVOTQEVOTQT 2290
D 2592 NOLLNQAQVNDQAGSKNDVDSAKTRSTQDIEQIOPHPQTKATGRHRLNEKAN---QOQS 2648
QY 2291 TVSSHVPSEAPQTHAQSKKPQ-----VAAQSQPQSN-----VQGGSPVRVQSPQTR 2337
```

```
D 2649 TIATHPNSTTEERQEAASAKLQAEVLKKAIAKIDKGTNDNDVEKTVVNGIAEIEINILPA-TT 2707
QY 2338 IRSTPSQLSPGQSQVOTTTSPQIPLOPHTSLQIPSOQPOSOPOVQSQSTOTLSSGQTL 2397
D 2708 VKDKAKADVNAERQK-----NQIINSNDEATTEELKVLASDNLNHVVET 2752
QY 2398 NOVSVSPSRPQLQIQOP-----QPOVIAVPPQLQQVQVLSQIQSOVVAQIAOQOS 2448
D 2753 NQAIEDAPDTNOVNVEKNKGIGITRDIQPLVVKKPTAK-----SKIESAV----- 2797
QY 2449 GVPOQIKLQPIQIQSSAVOTHOIQNVVTVVAASVOEQQLQVQOL-----RQOQKKKQ 2503
D 2798 -----EKKKTEINQTONATHDE---VREGLNQLNQHEKAKNDVNQSQTN 2839
QY 2504 QQIEINVTYPSKLLIKVEIILQKQVVMKHNVA-----IEHLKQKKSMTFAERE--- 2550
D 2840 QQVE---NAEONSLDQINFRPDSFKERNVAEIVKAQONKIDEIEQEFSAEQEKDNAL 2896
QY 2551 -----ENQRMIVCNQVMKY-----ILDKIDKEKQAA 2577
D 2897 QHLDEQVKEIINSINQANTDNEVDNAKTSGLNINITEYRPEYNNKKNAILKLYDVSDTQEA 2956
QY 2578 KKRREESVEQKRQONATKLSALLFKHKEQLRAEILKKRALLDKOLQIEVQBELK 2633
D 2957 IINGYPDATEDELOQAN-SKLNKILLDAKQOIGLAHTNNEV---DDIYNEVSQKMK 3008
```

RESULT 4

US-09-134-001C-3159

; Sequence 3159, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 3159

; LENGTH: 10182

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

; US-09-134-001C-3159

Query Match 2.6%; Score 391; DB 4; Length 10182;

Best Local Similarity 16.9%; Pred No. 1,3e-13;

Matches 500; Conservative 450; Mismatches 1114; Indels 894; Gaps 116;

QY 287 EDEWQCEVCAHKVPGVTDCAVEIQKNKPYIRHEPIGYDRS-----RRKYWFLNRRLLI 340

D 2666 KQLOQSIQPAASTDGTQDSTRNKNKQAEQAIQHANSVINNGDATSQINDAKNTV 2725

QY 341 EEDTENENKKTWYTSKVLAEIDCLDKDYW-----EALCKILIEEMREEIHRHMDITE 396

D 2726 EQAQDYVEAKSNLRADKSQLQSDYDTLNRDLTNDKPPASVRRYNEAISNTRKELDTAK 2785

QY 397 -DLTNKARGSKSF-----LAAANEELIESIRAKKGDIDNYS 433

D 2786 ADASSTLRNTNVEQVRDALNKINTQPKVNOQATLLOPKNNSELVQAKRLQDAVND 2845

QY 434 PETEKDKNETENDSKDAEKNREE--FEDQSEKSDDKTPDDDPQKSEPTVEV---- 487

D 2846 IPQTQGTQQTINNYNDKQREARLTSQAQVINDNGDATTOEITSEKSKVEQAMQALNTA 2905

QY 488 -----GDKG-----NSVSANLGDNTTATSETPSGRSPVGCSETPSSNMA 532

D 2906 KSNLRADKNELOFATYAKNLIENVSTN-GKKPASIRQVETAKARIONIN-----DAKNEA 2958

Db 991 PQQIGIPT-----SSLTQV-----VHSAGRRPIVSPVESRLRE 1024
QY 1421 GNVEPKVNNINKIIPENDIKSLAVKESAIRPFINGDVIMEDFNERNSETKSHLLSSDA 1480
Db 1025 SKVFP-----SEITDTVAASQAQSGNN--LSHSASSLSLQQAQFSELRAQMT 1070
QY 1481 EG-----NYRDSLEILPSTK---ESDSTOTTPSACPSNSVNOVEDMEIETSEVKV 1531
Db 1071 EGPNTAPPNFSHTGPTFPVPPFLSSLAGVPTTAAATAPVATSSPND--IST-----V 1124
QY 1532 TSSPITSEESNL-----SNDFIDENGLPKNNENNGESKRKTIVITEVTTMTSTVATES 1586
Db 1125 IQSEVIVPTTEGAGVATSTGVVTSGLPI---PPVSESPLVSSVSSITIPAVVSIIST 1181
QY 1587 KTVIKVEGDKQIVWVSTENCAKSTVTTTTTKLSTPSTGGSDVIISVKEQSKTIVTT 1646
Db 1182 SPSLOVPTSTSEIWSSTALYPSVTSATSA-----SAGGSTATPGKPPA--VVSQ 1231
QY 1647 TVTDSLTTTGTGLVTSMTVSKYSTDKVKLMKFRPKTRSGTALPSPYKRFYTKTKKS 1706
Db 1232 QAAGS--TTVGATLTSVSTTTSFSTASQLSI-----QLSSSTSTPTLAETVVSAAH-- 1281
QY 1707 IFVLPNDLKLARKGIREVPYFNYNNAKPAIDWPVPSRPRTFGIWRVRLQTVKSL-- 1764
Db 1282 -----SLDKTSHS---TTGLAFSLSAFSSSSSPG 1308
QY 1765 AGVSLMLRLLWASLRWDDMAAKVPPGGSGTRTETSEITEITTTIIRRDVGPYIRFEYC 1824
Db 1309 AGVSSVIS-----OPGG----- 1320
QY 1825 IRKIIICPIGYPETKPTP--TPQRKGLRSSALRKRPTPKQTGPV-----IETWVAEE- 1877
Db 1321 -----LHPLVIPSVIATPILPQAAGPTSTPLLPQVPSIPPLVQPVANVAVQOTLIHSQP 1376
QY 1878 -----ELE-----LWEIRAEERVE---KEKAQAEQOAKRLEEQ- 1910
Db 1377 QPALLPNQPHTHCPEVDSDTOPKAPGIDDIKTLEELKLSLFSHSSGGAQHASVLSLTS 1436
QY 1911 --KPTV---IATSTTPT---SSTTSTISPAQV---MVAPISGSVTTGKMLVTKVG 1958
Db 1437 VIESTVPTGPTTAVAPSCLLTSTSTCLPPTPLPLGTVALPVPVPTG-----QVS 1489
QY 1959 SPATVFOQKNHFOHFAIWWKQGSQVGVVQVQKVLGIIPSTGTSQOTFTSFQPTA 2018
Db 1490 TPVSTT-----TSQVKGPGTAPSKPPLTK 1512
QY 2019 TVTIRENTSGSGTNSOVITGPIRPGMTVIRTPLQ-----STLG 2061
Db 1513 APVLPVGTETLPAGTLSEQLPFP--GPSLTQSQQPLEDLDAQLRRLTSPXITVTSVAVG 1570
QY 2062 KAIRTPVMVOPGA--POQVMTQIIRQPVSTAVSA-----PNTVSTPGQKS--- 2107
Db 1571 PVSMAAPTATTEAGTQPKGVSOVKEGPVLATSSGAGVFKMGFRQVSVAAADGAQKEGKN 1630
QY 2108 -----LTSATSTNIQSSASQPPRQOGQVKLTHAQLTQITQHGNGOGLTVVIQGG 2159
Db 1631 SEDAKSVHFSSTSSSVLSSS-----PESTLVKPE----- 1662
QY 2160 GOTTGQLQLIPQGVTVLPGGQOLMOAAMPNGTVQRFLEFPTPLATTATTATTTT----- 2212
Db 1663 -----PNCITI--PGISSDVPESAH-----KTTASEAKSDGTQPTKVG 1698
QY 2213 -TTVSTTAAGTGQRQSK-----LSPQMQVHQDKTLPPAQSSTSSYPGAKAQP 2257
Db 1699 RFQVTTTANKVGRFSVSKTEDKITDKKEGPVASPPFMDLEQAVLP-----AVIPKKEP 1753
QY 2258 QTAQPSARPOPTQPO-----SPAOP----- 2278
Db 1754 ELSEPSHLNGPSDPEAAFLSRDVGSGSPHQLSSKSLPSQNLSSLSNFSNSYM 1813
QY 2279 -----EVO-----TOPEVOTQTTVSSHVP----- 2297
Db 1814 SSDNESDIEDDLKLELRRLRDLKHLKEIQDLSQRKHEIESLTKLGKVPVPAIIPPAAP 1873

QY 2298 ---SEAQTHAQSSKQVAA---QSOPQ---SNVOGQSPVRVQSPQSTRIRSTPSQLSPG 2349
Db 1874 LSGRRRPTKSKGSKSRSSSLGNKSPQLSGNLSQSAASVYLHPQQT-LHP--PGNIPES 1930
QY 2350 QOSQOVQTTTQOPPIQPHT-----SLQTPSQGQSQSQSQVOSSQTSLSSGQTLN 2398
Db 1931 GONQL-----LQPKPSPSSDNLXSAFTSDGAISVPSLSAP---QGGTSSNTV--GATVN 1981
QY 2399 QVSVSSPSRPQLQIQOQPQ 2417
Db 1982 S-----QAAQAP 1989

RESULT 6
US-09-854-856-16
; Sequence 16, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Waite, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; PRIOR FILING DATE: 2001-05-14
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1999
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1999)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-16

Query Match 2.5%; Score 380.5; DB 4; Length 1999;
Best Local Similarity 18.0%; Pred. No. 5.2e-14;
Matches 434; Conservative 297; Mismatches 757; Indels 921; Gaps 104;

QY 472 PDDPEQKSEETEVGDKGNSVANSGLDNTNATSE-----ETSPSEGRSPVCLSETP 526
Db 21 PAPAPKNGSSD--SSVGEK---LGAADAADVTGTEEYRRRRHMTMDKDSGAAATTTTTE 76
QY 527 -----DSS-----NMAEKKVASLPQDVPEEPNKTCESSNTSATTTSI- 564
Db 77 HRFRRSVICDSNATALEPLPLSLPQPIPAVPOQAPPEPHR--EETVTATATSOVA 134
QY 565 -QPNLENSNSSELNSQSASAKAADPEN--GERESHTPVISQEEIVGDTSEKSTGEL 621
Db 135 QQP-----PAAAPGEQAVAGPAPSTVPSTSKDRPVS-QPSLV-----GSK 175
QY 622 SESPGAGKAGSGSTRIITRLRNPDLSQLSKSQVAAAHEANKLFKEGKEVLVNSQGE 681
Db 176 EEPPIKSGSGGS-----AKEQEEKRSQ-----QDD 203
QY 682 ISRLSTKKEVIMKGNINNYFKLQGE---GKYRVVHNOYSTNSFA-----LN 724
Db 204 IELETK--AVGNSNDGRFLKFDIEIGRGSFKTVYKGLDTEVVEAWCELQDRKLTKE 261
QY 725 KQHREDHDKRRHLAH-----KFCLETPAGEFKWNGS-----VHGSKVLTI 764
Db 262 RQRFKEAEAMKGLQHPNIVRFYDWSWESTVKGKKCVILVTMLTSGTLKYLKRFKVMKI 321
QY 765 STLRLTITOLENNIPSSFFHPNW-----ASHRANW 794
Db 322 KVLRSWCRQKLGL--OFLHTRTPPIIHRDLKCDNIFITGTGSKVTKIGDLGLATLKRAFS 379

APPLICANT: Turner, C. Alexander Jr.
TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
FILE OF INVENTION: Encoding the Same
FILE REFERENCE: LEX-0178-USA
CURRENT APPLICATION NUMBER: US/09/854,856
CURRENT FILING DATE: 2001-05-14
PRIOR APPLICATION NUMBER: US 60/206,015
PRIOR FILING DATE: 2000-05-19
NUMBER OF SEQ ID NOS: 64
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 2108
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(2108)
OTHER INFORMATION: Xaa - Any Amino Acid
US-09-854-856-30

Query Match 2.5%; Score 380.5; DB 4; Length 2108;
Best Local Similarity 17.9%; Pred. No. 5.6e-14;
Matches 430; Conservative 300; Mismatches 751; Indels 925; Gaps 105;

QY 472 PDDDPQGGSEETEVGDGKNSVANSALGNDTWTNATSE-----ETSPSEGRSPVGCLESTP 526
D 21 PAPAPKNGSSD--SSVGEK---LGAADAAVATGRTTEYRRRRHTMDKDRGAAATTTTTE 76
QY 527 -----DSS-----NMAEKKVASELPQDVPPEPNKTCESNTSATTSI- 564
D 77 HRFRRSVTCDSNATLELPLSLPQPSIPAAPVQSPAPPEPHR--EETVTATATASQA 134
QY 565 -QPNLENSSSSLSNLSQESAKAADPEN--GERESHTPVSTQEEIVGDFTEKSTGEL 621
D 135 QQP-----PAAAPGEQAVAGPAPSTVPSPSTSKDRPV- QPSLV-----GSK 175
QY 622 SESPGAGKASGSTRITRLRPNDPSKLSQKSOQVAAAHAANKLFKEGKVLVNSQGE 681
D 176 EEPPEARSGGGS-----AKEPQERSQ-----QQDD 203
QY 682 ISRLSTKKEVIMKGINNYFKLQGE--GKYRVYHNOYSTNSFA-----LN 724
D 204 IELETK--AVGNSDGRFLKFDIEIGRSGFKVYKGLDTETTVFVAVCELQDRKLTKE 261
QY 725 KHQREDHDKRRHLAH-----KFCLTAPAGEFKWNGS-----VHGSKVLTI 764
D 262 RQRFKEEAEMKLQHPNIVRFYDSWESTVKGKKIVLVTMLTSGTLKYLKRFKVMKI 321
QY 765 STLRLITQLENNIPSPFHPNW-----ASHRANW 794
D 322 KVLRSWCRLKGL--QFLHTRTPPIIHRDLKCDNIFITGPTGSKIGDLGLATLKRASF 379
QY 795 IKAV-----QMSCKPRE-----FALALATLECA-----V 818
D 380 AKSVIGTPEMAPEMEYKDESVDYAFGCMLEMATSEYPYSECONAAQIYRVRTSGV 439
QY 819 KPV-----VMLPIWRE-----FLGHTRLHRMTSI-----EREKEK-- 849
D 440 KPASFQKVAIPEVKEIIEGCIQKNDERYSIKDLLNHFQEEYGVYVELAEDDGEKIA 499
QY 850 -----VKKEKKOEETEMQOATWYKTPPVKHQVWVKQGEYRVTYG---GGWSW 897
D 500 IKLWLRIEDIKLKGYKQNEA-----IEFSPDLERDVPEDVAQEMWESGYCEGDHKT 553
QY 898 ISKTHYVRFVKPLPGNTNRYKSLGKTKNNMDENMDESCKRCSRPKKIKLEPSEKD 957
D 554 MAK-----IKDRVSLIKRKEQQLVREGEKK-----KQESLSLKQ 591
QY 958 EYKGSAAKAGQONEMDISKITEKKDQDVKLELSDSDSKPKEEPMEDDDMKTESHVNC 1017
D 592 QVEQSSASQ-----TGKQLPSASTGIP-----TASTTSA 621
QY 1018 QESSQVDVNVNSEGFHLRTSYKKKTKSKLDGLLERRIKOFTLEEKORLEKIKLEGKIG 1077

Db 622 SVSTQVEPEPEADQHQLOYOQPSISVLSDGTVD----- 656
QY 1078 IGTSTNSSKNLSESPVITKAKGECQSDMRQSQSNANNDQPEDLIQGCQSQSDSSVLRM 1137
Db 657 -----SQGSSSVFTESRV-----SSQIVSYGSGHQEAHSTGTVPGHIPSTVQAO 701
QY 1138 SDPSHTNKLKPKDRVLDDVYSIRSPETKCPKQNSIENDI-----EEKYSDLASRQEP 1190
D 702 SQP-----HGVYPPSSVQOGI-----QQTAPQOQTVQVLSQSTSSSEATTAAQPVSQPAP 752
QY 1191 T-----KSKTKGNDFFDDSKLASADDIGLICKNKKPLQEEESDTIVSSSKSALHS 1242
D 753 QVLPQVSAGKQSTQG-----VSQVAPAEPAV-----AQP--QATQPTTLASSVDSAHS 799
QY 1243 SVPKSTNDRDATPLSRAMDFEGK-----LGCDSSENSTL 1276
D 800 DVASGMSDGNENVPSSSGRHEGRTTKRHYKRSVRSRHEKTSRPKRLILNVSNKGRVV 859
QY 1277 ENSSDT-----VSIQ-----DSSEE-----DMIVONSNEISEQFRTREODVEYLEP 1318
D 860 ECQLETHNRKMWTFKFDLDGDNPEEIATIMVNDNFILAIERESFVDQVR---EIEKADE 916
QY 1319 LKCELYSGESTGNCBDRLPVKGTBANGKPKSQOKKLEER---PVNKCSDQIKLKNNTDKK 1375
D 917 MLSEDYSVEPEGD--QGLESLOQKDDYGFSGOKLEGEFKQIPASSMPOQIGIPT----- 970
QY 1376 NENRESEKKQORTSTFOINGKDNKPKIYLKGE--CLKEISESRVSVGNVEPKVNNINKI 1433
D 971 -----SSLTQV-----VHSAGRFRFVSPVPSKLRKESKVPF----- 1001
QY 1434 IPENDIKSLTVKESAIRPFINGDIVMEDFNERNSSSETKSHLLSSSDAEG-----NYRDS 1487
D 1002 ---SEITDVTAASTAQSPGN---LHSASSLISLQQAFLSELRAQMTGEPNTAPPNESH 1055
QY 1488 LETLPSTK---ESDSTQTTTPSACPSNVSNOVEDMEIETSEVKKTVSPITSEESNL 1544
D 1056 GPTFPVVPPLSLIAGVPTTAAATAPVPATSSPND--ISTS-----VIQSEVTVPTBEG 1109
QY 1545 ---SNDFTDENGLPINKNENGVESKRKTVITEVTMTSTVATESKTKVKEGDKQT 1599
D 1110 AGVATSGVTVSGGLPI---PPVSESPVLSVVSSITIPAVVISITSPSLQVPTSTSEI 1166
QY 1600 VVSTSTENCAKSTVTTTTVTKLSTPSTGGSDIISVKEQSKTVVTVTTVDTSLTTGGTL 1659
D 1167 VWSSTALYPSVTVSATSA-----SAGSGTATPGPKPPA--VVSQAAAGS--TTVGATL 1215
QY 1660 VTSMTVSKYSTDKVKLKFSPKPKTRSGTALPSYKFKVTSTKKSIFVLDPNDLKLKLA 1719
D 1216 TSVSTTTSPSTASQLSI-----QLSSSTSTPTLAETVVSAAH----- 1253
QY 1720 RKGIREVPYFNNAKPAIDWPPSPRPTFGITWRYLQTVKSL--AGVSLMLRLWLAS 1777
D 1254 -----SLDKTSHSS--TTGLAFSLASPSSSSPGAGVSSYIS----- 1288
QY 1778 LRWDDMAAKVPPGGGSTRTETSETIITTEIIRRDVGPYIRFEYCIKRIICPIGVPET 1837
D 1289 -----QPGG-----LHPLVIPSV 1301
QY 1838 PKETP--TPQKRGRLRSALRPKRPETPKQTGPV-----IETWVAEE----- 1877
D 1302 IASTPILPQAAGTSTPPLPQVPSIPPLQVPAVNPVAVQOQLIHSQPPALLNPQPHTHC 1361
QY 1878 -ELE-----LWEIRAFARVE---KEKAQAVEQQAARLEQ---KPTV---IATS 1918
D 1362 PEVSDTQPKAPIDDIKLTLEKLSLSESSGGAQHASVLSLTSVLIVSTVTPGIPPT 1421
QY 1919 TTSPT---SSTTSTISPAQKV---MVAPISGVTTTGTCKMVLTKVGSFATVTFQOKNKF 1971
D 1422 AVAPSKLLTSTTCLPPTNLPLGTVALPVPVTPG-----QVSTPVTST----- 1467
QY 1972 HQTFAWVKQGSNGSVVQVKVLGIIPSTSQOTFTSFQPRATVATVIRPNWTSGGG 2031

Wed Sep 24 09:18:12 2003

1468 -----TSGVKGTAPSKPPLTKAPVLVGTETLPG 1497
2032 TTSNOVITGQIRPGMTVIRTPLOQ-----STLGKAIIRFPVMVOPG 2074
1498 TLPSPQLPFP--GPSLTQSOQPLEDLAQLRRLTSPXITVTSAVGPVSMAPTAITEA 1555
2075 A--POQVMTQIRGPVSTAVSA-----PNTVSTPGQKS-----LTSAT 2112
1556 GTPQKGVSQVKEGVPVLTATSGAGVFKMGRFQVVAADGAGKCKNKSEDAKS VHFESST 1615
2113 STSNIOSSASQPPRPOQGOVLTMAQLTQLTQGHGNOGLTVIOGOGQTTGQLQLIPQG 2172
1616 SESSVLSSSS-----PESTLVPE-----PNG 1637
2173 VVVLPGQOQLMQAAMPNGTVQRFLETPLATATTATTT-----TTVSTTAAGTGE 2224
1638 ITI-PGSSDVPEAH-----KTTASEAKSDTQPTKVGRFQVTTTANKVGR 1683
2225 QRQSK-----LSPQMVGHDKTLPPAQSSSVGPAKAOPOTAOPSARPOT 2270
1684 FVSKTEKIDTKKEGVPVSPFFMDLEQAVLP-----AVIPKKEKPELSEPHLNGPSS 1738
2271 QPQ-----SPAQP-----2278
1739 DPEAAFLSRVDGSGSPHSPHQLSSKSLPSQNLSQLSLSNFSSSYMSDDNESDIEDDL 1798
2279 -----EVO-----TOPEVQOTTVSSHVP-----SEAQPTHAQS 2307
1799 KLELRLRLDKHLEIQDLSROKHEIESLYTKLGKVPVPAVIPPAPLPGRRRRPTKSG 1858
2308 SKPQVAAA--QSQPO--SNVOGQSPRVQSPSOTRIRPSTPSQLSPGQSQVQTTTSQP1 2362
1859 SKSSRSSLGNKSPQLSGNLSCQSAASVLPQOT-LHP--PGNIPESQNL-----LQPL 1911
2363 PIQPH-----SLQIPSGQPOQPOVQSGSTQTLSSGQTLNOVSSVSPRPQLO 2411
1912 KPSPSSDNLVSAFTSDGALSVPVLSAP---GQGTSTNTV--GATVNS-----Q 1955
2412 IQQPOP 2417
1956 AAQAQP 1961
RESULT 8
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 1800
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11
Query Match 2.5%; Score 380; DB 4; Length 3878;
Best Local Similarity 16.8%; Pred. No. 1.4e-13;
Matches 488; Conservative 476; Mismatches 1025; Indels 922; Gaps 115;
305 DCVAEI-----QKNKP--YIRHEPIGYDRSRKWTW-----FLNRRLLIE 341
125 DCSSEVNGCSFVMRTGKPTLLRREEFGVDVDSYEQGAQDSPTHLEMSELAGKQHEIE 184
342 EDTENENEKKIWIYSTKVQLAELIDCL-----DKDYWEAELCKIL 381

185 ELNRELEMRVTVGTGEOLOQOEFAAIKOROGIITQLTANLQOARREKDETMREFLELT 244
382 EEMRE-EIH-RHMDITEDLTNKARGSNKSFLEAANEIELESIRAKKGDIDNVKSPETEK 439
245 EQSKLQIQOQLOASETLRNSTHSTAAADLQAKQOILTH-----QQ 287
440 DKNETENDSKDAKNEEFEDQSLKDDDKTDDPDEQKSEEPTEVGDGKNSVANLG 499
288 QLEQDHLLEDYQKKKEDFTMQI--SFLQEKIKVYEMEQDKVE-----329
500 DNTNATSETSETSEGRSPVGCCLSETPDDSNMAEKKVASLPDQVPEEPNKTCESNTSA 559
330 ---NSNKEIEQKE-----TIIIEELNTKIIIEEKKTIIELKDKLT 365
560 TTTS-----IOPNLENSSSSELNSSQSESAKAADD-----PENGERSHTPSV 603
366 TADKLLGELOEQIVQKQOEIKNKKLELTNSKQEROSSEIKQLMGTVELOKRNHKDSQ 425
604 IQEEIVGDTSEKSTGSELSPGAGSGSTRIITRLRNPDSKLSQKSOQVAAAHEA 663
426 FETDIVORMEOE-----TORKLEQLR--AELDEMYGQIV-----458
664 NKLFKEKEVILVNSQGEISRLSTKKEVIMKGNINNVKLGQEGKYRVYHNOYSTNSFAL 723
459 ---OMQELIIROHMAQMEEMKTRHKGEMENALRSYSNI-----TVNEDQIKLMNVAI 507
724 NKQHRR-EDHDKRRH-----LAHKFCLTPAGEFKWNGSVHGSVKVLTITLRLTITQ 773
508 NELNKLQDINSQEKLEKEELGLILEEKCALQ-----RQ 541
774 LENNIPSSFPHNWAHRAHWIKAVQCMSPREFALAILCAVAKPVVMLPTWRFLGH 833
542 LEDLVEELSFSEIQIARQTI--AEQESKLNHAHSLSTVEDLKAEIVSASRSKELEL 599
834 TRLHRMFS-----IEREKEKVKKEKEEBEETMQOATWVYTPFPVKHVMQKGE---E 886
600 KHEAEVNYKTKLEMEKEKNAVLDRMAESQAELERLTOLLFSHEEELSKLEDLEIE 659
887 YRVTYGGGWSWISKTHVYRVPKLPNGTNVNYKSLGCTKNNDENMD--ESDKRKSRS 944
660 HRIN-----IEKLDNLGIHYKQOIDLQNLQNSQKITEKQFQKDNLTIK 703
945 PKKIKIEPDSKDEKVGSDAAKAD-----QENEMDISKITEK-----DQDYKEL--- 989
704 QNLILEISKLDLQOQSLVNSKSEMTLOINELQKEIILRQEEKEKGTLEQEOELQLK 763
990 -----LSDSDKPKCKEENVEVDDMMKTESHNVCQSSQVVDVNV 1028
764 TELLEKQMKENDLQEKFAQLEAE-NSILKDEKKTLEDMLKIHTPVSQEE-----813
1029 SEGPHLRTSYKKTKSSKLDGLLERRIKQFTLEEKORLEK--IKLEGKIGKGTSTNSS 1086
814 -----RLIFLDSIKSKSDSVWEKEI--EIIIEENEDLKQOCIQLNIEIEKQRTFSFAE 866
1087 KNL-----SESPVITKAKEGQSDSMROE---QSPNANDQPEDLIQGCQSQSSV 1134
867 KNFEVNYQLOEYACLLKVKDDLEDSSKNKQOEKYSKLLKALNEE-----911
1135 LRMSDPSHHTNKLYPKDRVLDVSRSPET-----KCPKQNSIENDI 1176
912 LHLQRINPTTVKM--KSSVFDKDTFVAETLEMGEVVEKDTTELMKELVTKREKLE--L 967
1177 EEKVSDDLASRGQEPKTSKTKGNDFFIDDSKLASADDIGTLICKNNKPLIQEESDTIVSSS 1236
968 SORLSDL-----EQLKQKHGEISFLNEEVKSLQOE-----KEQVSLRCLRELEIINHN 1016
1237 KSALHSSVPKSTNDRDATPLSRAMDFEGLGCDSESNSTLNSSTVSIQSDSEEDMIVQ 1296
1017 RA-----ENVQSDCTQVSSIL-----DGVVTMTSRGAGSVS 1048
1297 NSNESISEQFRTREQDVEVLEPLKCLYSGESTGNC-----DRLP--VKGTEANGKPSQ 1351
1049 KVNKSFGEESKINVEDKVSFE-----NMTVGEESKQELILDHLPSTKESLRATQPSN 1104

Wed Sep 24 09:18:12 2003

SOFTWARE: PatentIn Release #1.0, Version #1.30

DO NOT WRITE IN THESE SPACES

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08

FILING DATE: 16-SEP
CLASSIFICATION: A35

CLASSIFICATION: 433
ATTORNEY/AGENT INFORMATION:

NAME: Frommer Esq., William S.

REGISTRATION NUMBER: 25,506

REFERENCE/DOCKET NUMBER: 454312-2460

TELEPHONE: (212) 840-3333

TELEPHONE: (212) 840-3535
TELEFAX: (212) 840-0712

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 8991 amino acids
TYPE: amino acid

TYPE: AMINO ACID
STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: amine

8-714-741-32

2.58; Score 378; DB 4; Length 8991;

Query Match 2.5%; Score 378; DB 4; Length 8991;
Best Local Similarity 16.7%; Pred. No. 6.4e-13;
Matches 592; Conservative 475; Mismatches 1390; Indels 1082; Gaps 132;

Qy	1	MVSEEEEDGDA--RETQDSEDEED---EMEEDDDSDYPREMEDDDDDADVCYCTESSF	55
Db	1545	VASQSAEKDYDAAVEKSKAAEDELBEAAQAQRYDEQDKSEENKETEAS-----	1597
Qy	56	RSHSTYSSTPGRRKPRVHRPRSPILBEKDIPPLEFPKSSEDLM--VPNEHIMNVAIYEV	113
Db	1598	-----ERQQAATLKHLESXEFLNYFQDHRMNNKKM--I	1630
Qy	114	LRNFGVTLRLSPRFEDFCAALYSQBOCTLMAEHVVLLKAVLREE--DTSNTFTGPDALK	172
Db	1631	LTSLASVAILG-----AGLVASPTVVRAEEAPVASQSAEKDYDTAK-----R	1674
Qy	173	DSVNSTLYFDGMTWPEVL-----RVYCESDKYHHVLPYQEA--EDYPGYPVENKIKVLQ	226
Db	1675	DAENAKKALEAKPAQEKYADYQRRIEEKAAKETHASLEQOEAKNDY-----QLKLKK	1727
Qy	227	FLVDQFTTNIAREELMSEGVIQDDHCRVCHKLGDLCCETCSAVYHLECVKCPLEVP	286
Db	1728	YLDGRNLSNSVLKKEMEEAKKDKKPAEFNKI-----RREIVVPNPOELE	1774
Qy	287	EDWQCEVCVAHKVPGVITDCVAEIQKNKPVIRHEPIGYDRSRRKYWFNLRRLLIIEEDTEN	346
Db	1775	MARKSEVAKT--KESGLVKRVEEAKKVTEARPK--LDAERAEVVLQAIAM-----	1824
Qy	347	ENEKKIWYS-----TKVQLAEL-----IDCLDKDYEAEE-----LCKIL	381
Db	1825	--NKKKMILTSLASVAAILGAGLVASPTTVRAEEAPVASQSAEKDYDTAKRDAENAKKAL	1883
Qy	382	EMEEREIHRHMDITEDLTNKARG-----SFKSFLAAA	413
Db	1884	EEAKRAQEKYADYQRRIEEKAAKETHASLEQOEAKNDYQLKLKYLDRNLSNSVLKKE	1943
Qy	414	NEETLESIRAKKGIDN-----YKSPETEEDK	440
Db	1944	MEEAEEKDKKQAGLNNKKMILTSLASVAILGAGLVTSQPTVLRAEESPVASQSAEKD	2003
Qy	441	KNETENDSKDAEKNREE-----PEDOSLEKD-----	466
Db	2004	YDAAKRAENAKKALEEAKRAQEKYADYQRIEKAKEQASLEQQEANKDYQLKLKKY	2063
Qy	467	SDDTTPDDPBQKSEEPTEVGDKGNSVAN--LGDNTTNTATSEETSPSEGRSPV-----	519
Db	2064	LDGRNLSNSVLKKEMEEAKKDKKQAEFNKIRREIVVPNPOELEMARRESEVVKAKES	2123
Qy	520	GCLSETPDSSNAEKVY-----ASEL---PDQYPEPNKTCESNN-----TSA	559
Db	2124	GLVRKVEE---AEKKVTEARQKLDAAERAKEVVOPTRVENEVHKLXQMKNNKKMILYSL	2173

Matches 422; Conservative 290; Mismatches 755; Indels 844; Gaps 96;

Qy	508	EETSPSEGRSPVGCISTPDSNMAEKKVASELPQDPEEPNKTCESSNTSATTSIQPN	567	1409	CLKEISESRVSGNVEPKVNNINKIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSS	1468
Dd	62	EETVATATSOVA--QPPAAAACEQAVAGAPASTVP-----SSTSKDRPVSQPS	110	953	IVSPVPESRLRESKVPF-----SEITDTVAASTAQSPGMN---LSHSASSLSLQ	998
Qy	568	LENSNSSSELNSESOSAKAADDPENGERESHTPVSIQEEITVG-----DFTSEKSTG	619	1469	ETSKHLLSSSDAEG-----NYRDSLETLPSTK---ESDSTQTTTPSPASCPESNVNOVE	1519
Dd	111	LVSKEEPPPARSGGGSAKEPQERSQQODDIELETRAAGVNGDRFLKFDIEIGRG	170	999	QAFSELRAQMTGEPNTPAPPNFSGTPTFPVPPFLSSIAAGVPTTAATAAPVATPPATPPN	1058
Qy	620	ELSESPGAGKAGSSTRI--ITRLRNPDSKLSQKASQOQVAAAHEANKLFREGKEVLVNS	678	1520	DMEIETSEVKVTSPITSEESNL-----SNDPIDENGLPINKNNVNGESKRTKIVITE	1574
Dd	171	SFKT---VYKGLDTEVAVWCELODRKLTSEOR-----FKEEAEMLRGLQ	216	1059	D--ISTS---VIOSEVTVPTTEGIAVATISGVTSGGLPI---PPVSESPVLSVVSS	1109
Qy	679	OGETSRLSTKKEVIMKGN--TNNYFKLGOEKKYRVYHNOYSTNSFALNKHOHRD----	731	1575	VTTMTSVATESKTVIKVEKGDQTVVSVSTENCASKATVTTTTTTTVTKLSTPSTGGSDII	1634
Dd	217	HPNIVRFVDSVESVKGKCVILVTELMTSTLTKYLRKPKVMKIVLRSHCRILKGLQ	276	1110	ITIPAVVISISTSPSLQVPTSTSEIVVSTALYPSVTSATSA-----SAGGSTATP	1161
Qy	732	--HDKRRHLAHK-----FCLTPAGEPKWNGSVHGSVKVLFISTRLTITOLENNIPSSF	782	1635	SVKQSKTVVTTTVDLSLTITGGTLTVMTVSKESYSTRDKVLMKFRPKKTRSGTALPS	1694
Dd	277	FLHTTPPIIHRDLKCDNIFITGTGSVKIGD-----LGLATLK-----	315	1162	GPKPPA--VVSQQAAGS--TTVGATLTSVSTTSPSTASQLSI-----OLSSSTSTPT	1211
Qy	793	PHPNWASHRAWIKAV-----QMSCKPRE-----FALALAILICA-----	817	1695	YRKFVTKSTKKSIFVLPNDOLKILARKGGIREVFNYNAPKALDIWYPSPRPTFGITW	1754
Dd	316	-----RASFAKSVIGTPEFMAPEMEYEEKYSDVDVYAFGCMCMLEMATSEYPYSECQ	367	1212	LAETVVVSAH-----SNDKTSHSS-----TTGLAF	1236
Qy	818	-----VKPV-----VMLPIWRE-----FLGHLRHMSTI--	842	1755	RYRLQTVKSL--AGVSLMLRLMLASLRWDDMAAKVPPGGSGSTRTETSETETITTEILKRR	1812
Dd	368	AAQIYRRVTSVKPASFDKVAIPVEKELIEGCIQONKDERYSIKDLLNHAFFQBEETGVRV	427	1237	SLSAPSSSSSPGAGVSVIS-----QPGG-----	1260
Qy	843	-----EREKEK-----VKKEKKQEEETMQQATWYKVTFFPVKHQWVKOGGEYR	888	1813	DVGPGYGFEXCIRKIICPIGVETPKETP--TPORKGLRSSALRPKRPETPKQGPV---	1868
Dd	428	ELAEEDDEKTAIKLWLRIEDIKKLKGKDYKNEA-----IEFSFDLERDVPEDVAQEMV	481	1261	-----LHPLVIVSIASTPILPQAAGTSTPLLPQVPSIPPLVQPVANV	1304
Qy	889	VTGY---CGWWSIKTHYRVFKLPNGTNYNRYKSLBGTKNNDENMDSDKRCSSRP	945	1869	--LIETWVAEP-----ELE-----LWEIRAFARVE---KEKAQAV	1899
Dd	482	ESGYVEGDHKTMAKA-----IKDRVSLIKRREORQLVREGEKK-----	522	1305	PAVOOTLIHSQPOALLPNQPHTHCPEDVSDTQPKAPCIDDIKTLEELRSLFSEHSSG	1364
Qy	946	KKIETPEDEKDEKVGDAAGQONEMDISKITEKKDQDKVLELSDSDRCKPEEPMEV	1005	1900	EQAKKRLLEOQ---KPTV---IATSTSTPT---SSTSTSTSPAKV---MVAIFSSVT	1946
Dd	523	---KQEESSLKQOVEQSASQ-----TGIKLPASASTGIP-----	554	1365	AQHASVSLTSLVISTESTVTPGTPPTAVAPSKLLSTTSTCLPTNLPLGTVALPVPVPT	1424
Qy	1006	DDMKTESHVNCQESSQVDVNVNVEGFHLRTSYKKKTKSKLDGLLERRIKOFTLEEKOR	1065	1947	TGTRKVLTTKVSATVTFQONKNFHQTFATWVAQGSNGVGVQQKVLGIIPSSTGTS	2006
Dd	555	-----TASTSASVSTQVEPEPEADQHQQLQYQOPSISVLSGTVD-----	596	1425	PG-----QVSTPVSTT-----TSQVK	1440
Qy	1066	LEKILLEGIGIKGTSTNSKNLSESPVITKAKEGCOSD-----SMROEQS	1112	2007	QOTFTSFOPRATVITIRENTSGSGTTSNSQVITGPQIRPGMTVIRIPLQO-----	2057
Dd	597	-----SGQSSVFTESRVSSQQTV-----SYGSQHEQAHSHTGTVPGHIPSTVQAQS	642	1441	PGTAPSKPPLTKAPVLPVGTLPAGTLPSEQLPPFP--GPSLTQSQOPLDLDQAQRRTL	1498
Qy	1113	PNANNDQPEDLTQGCQSDSSVLRMSDPSTHTNKLYPRKDRVLDVYSIRPETKCPKQNSI	1172	2058	-----STLGKALIRTPVMVQPGA--PQVMTQIIRGQVPVSTAVSA-----PNTV	2099
Dd	643	QPHGVYPPSSVAQOGSQPSSSLTGVSSQPIQHPQOQ-----QGIQOTAPPOQTV	695	1499	SPEXITVTSVAGVPVSMAPTAITAGTQPKGVQVQKGPVLTSSGAGVFKMGRFQVSV	1558
Qy	1173	ENDI-----BEKVSDILASRGQPT-----KSKTKGNDFFIDDSKLASADDIGILI	1217	2100	SSTPGOKS-----LTSATSTSIQSSASOPPRQOQGVKLTMAQLTQLTOGHG	2147
Dd	696	QYSLSQSTSTSEATTAAQVPVQAPQVLPQVFSAGKQSTQG-----VSQVAPAEVAV--	747	1559	AADGAQKGGKNSDEKASVHFSESTSESVLSSSS---PESTLVKPE-----	1602
Qy	1218	CNKKPLIQEEDTIVSSKSAHALHSSVPKSTNDRDATPLSRAMDFEGK-----	1265	2148	GNQGLTVVIOGQGTGOLQIIPQGVTVLPQGOOLMOAAMPNGTVQVRFLETPPLATTAT	2207
Dd	748	---AQP---QATOPTTLASVDASHDVAAGSDGNGENVPSGSGRHEGRTTKRHKRYSRS	802	1603	-----PNGITI--PGISSDVPESA-----KTTASE	1626
Qy	1266	-----LGCDSSENSTLENSSDT-----VSIQ-----DSSEE-----DM	1293	2208	ASTTT-----TTVSTTAAGTGEOQROSK-----LSPOQMVHQDQLPPA	2245
Dd	803	RSRHEKTSRPLRLNVSNKGDVVEQCOLETHNRKMWTFKFDLGDGNPEEITATIMVNDNF	862	1627	AKSDTGQPTKVGRFOVTTTANKVGRFSVSKEDKITDTKKGGPVASPPFMDLEQVLP--	1684
Qy	1294	IVQNESISEQFRREQDQVEVLEPLKCELVSGETGNCEDRLPKVGTAEANGKKPSQOKK	1353	2246	QSSSVGPAKAPQAPQAPQAPQAPQ-----SPAOP-----	2278
Dd	863	ILAIERESVDQVR---ETIIEKADEMLSEDVSVEPEGD--QGLLESQGDYDGFSGSKLE	918	1695	---AVIPKKEPELSEPHLNGPSSDPEAFLSRDVGSGSPHSPHQLSSKSLPSONLS	1741
Qy	1354	LEER---PVNKCSDQILKANTDDKKNNENRESEKKGQRTSTFOINGKDNKPKIYLKGE--	1408	2279	-----EVQ-----TQPEVQOTQTVSSH	2295
Dd	919	GEFKQIPASSMPQIQIGIPT-----SSLTQV-----VHSAGRRF	952	1742	QSLSNSFNSSYSSSDNESDIEDDLKLELRLRLRDLKHLKLEIODLQSRQKRIESTLYTKLGK	1801

Qy 2296 VP-----SEAQTHAQSXPQVAA---SQSQPQ---SNVQSQSPVRVQSPQSTR 2337
Dy 1802 VPPAVIIPAPLPSGRRRRPTKSGKSSRSRSLGNKSPQLSGNLQSQSAASVLPQQT- 1860
Qy 2338 IRPSPQSPGOQSOVQTTTSPQPIQPH-----SLQIPSQPOQPOQVQS 2386
Dy 1861 LHP--PGNIPESGONQL-----LQPLKPSSDNLYSAFTSDGALSVPSLAP---GOGTS 1911
Qy 2387 STQTLSSGOTLNQVSVSSPSRPLQIQPOP 2417
Dy 1912 STNTV--GATVNS-----QAAQAQP 1929

RESULT 11
US-09-854-856-32
; Sequence 32, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 1971
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1971)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-32

Query Match 2.58; Score 374.5; DB 4; Length 1971;
Best Local Similarity 18.08; Pred. No. 1.2e-19;
Matches 431; Conservative 294; Mismatches 748; Indels 923; Gaps 104;

Qy 472 PDDPEQKGSEETEVGDKGNSVANSGLDNTNATSE-----ETSPSEGRSPVGCLETP 526
Dy 21 PAPAPKNGSSSD-SSVGEK---LGAADAADAVTGTEYRRRRHTMDKDSRGAATITTE 76
Qy 527 -----DSS-----NMAEKKVASSELPODVPPEPNKTCESNTSATTTSI- 564
Dy 77 HRFRRFRSVIDSNATALEPLPLSLPQPSIPAAPVQSPAPPEPHER--EETVTATATSQVA 134
Qy 565 -QPNLENSNSSELNSQSSAKAADPEN--GERESHTPVTSQEEVLGVFTSEKSTGEL 621
Dy 135 QQP-----PAAAPGQAVAGPAPSTVPSTSKDREVS-QPSLV-----GSK 175
Qy 622 SESPGAGKAGSGSTRIITRLRNPDSKLSQLKSQVAAAHAEANKLFKEGKVLVNSQGE 681
Dy 176 EEPFARSGGGS-----AKEQEERSQ-----QQDD 203
Qy 682 ISRLSTKKEVIMKGNINNYFKLQGE---GKYRVYHNOYSTNSFA-----LN 724
Dy 204 IEELKTK--AVGMSNDGRFLKFDIEIGRGSEFKTVYKGLDTEITVEVAWCELQDRKLTKE 261
Qy 725 KHQHREDHRRHLAH-----KFLTPAGEFKWNGS-----VHGSKVLT 764
Dy 262 RQRKEEAEMKGLQHPNIVRFYDSWESTVKGKKCIVLVTLMTSGTLTKYLKRFKVMKI 321
Qy 765 STLRLTITOLENNIPSSFFHPNW-----ASHRANW 794
Dy 322 KVLRSWCRLKGL--QFLHTRPTPIIHRDLKCDNIPITGPTGSVKTGDLGLATLKRASF 379

Qy 795 IKAV-----QMCSPRE-----FALALAILLECA-----V 818
Dy 380 AKSVIGTPEFMAPENMEKYDESVDVYAFGCMCMLEMATSEYPYSECQNAAIYRRVTSGV 439
Qy 819 KPV-----VMLPIWRE-----FLGHTRLHRMTSI-----EREKEK-- 849
Dy 440 KPASFDKVAIPEVKEIIEGCIHQKNDERYSTKDLNLHAFQOEETGVRVELAEDEGKIA 499
Qy 850 -----VKKKEKKOBEETMOQATWVKYTFPVKQVWKQKQEEYRVTY--GGWSW 897
Dy 500 IKLWLRIEDIKLLGKYKNEA-----IEPSFLERDVPEDVAQEMVESYVCEGDHKT 553
Qy 898 ISKTHVYRFVPLPGNTNVNVRKSLGKTKNMNDMDSDKRRKCSRSRPPKIKIIPDSEKD 957
Dy 554 MAK-----IKDRVSLIKRKRQRLVREEQEK-----KQESSLUKQ 591
Qy 958 EVKGSDAAGADQONEMDISKITEKKDQVKKLLSDSDKPKCKEPMVEVDDDMKTESHVNC 1017
Dy 592 QVEQSSASQ-----TGIKOLPSASTGIP-----TASTTSA 621
Qy 1018 QESQVDVNVVSEGFHLRTSYKKTTSKSLDGLLERRIKQFTLEEKORLEKIKLEGGIKG 1077
Dy 622 SVSTQVEPEPEADHQOQLOQOQPSISVLSGTVD----- 656
Qy 1078 IGTSTNSKNLSESPVITRAKEGCSQSDSMRQSPNANNDDPDLIQGCSQSDSSVLRM 1137
Dy 657 ----SGQSSSVFTESRV-----SSQOTVSYGSHQEAHSTGTPVGHIPSTVQAQ 701
Qy 1138 SDPSTTNKLYPKDRVLDVYSIRSPETKCPQKQNSIENDI-----EKKVSDLASRQEP 1190
Dy 702 SQP-----HGVPYPPSSVQOGI-----QOTAPPQOTVQVYLSQTSSTSEATTAPQVSPQAP 752
Qy 1191 T-----KSKTKGNDFFIDDSKLASADDITGLICKNKRLQIEESDTIVSSSKSALHS 1242
Dy 753 QVLPQVSAGKQSTOG-----VSQVAPAEPAV-----AQP--QATQPTLIASSVDSHS 799
Qy 1243 SVPKSTNDRDATPLSRAMDFEGK-----LCGDSSESNTL 1276
Dy 800 DVASGMSDGNENVPSSSRHREKTRKRYKSVRSRSHKTSRPKRLILNVNSKDRVV 859
Qy 1277 ENSSDT-----VSIO-----DSSEE-----DMIVQNSNESISEQFRTQEDQVLEP 1318
Dy 860 ECQLETHNRKMTVKFDLGDGNPEEIATIMVNNDFILAIERESFQDVR-----ETIEKADE 916
Qy 1319 LKCELVSGESTGNCEDRLPVKTEANGKPSQOKKLEER---PVNCKSDQIKLNTTDDK 1375
Dy 917 MISEDVSEVEDG--QGLSELOGKDDYGFSGSQLEGEKQPIPASSMPQOIGITP----- 970
Qy 1376 NENRESEKKQRTSTFTQINGKDNKPIYKGE--CLKEISESRVSGNVEPKVNNINKI 1433
Dy 971 -----SSLTQV-----VHSAGRRFIVSPVPSRLRESKVFP----- 1001
Qy 1434 IPENDIKSLVKEAISRPFINGDIVMEDFNERNSETKSHLLSSSDAEG-----NYRDS 1487
Dy 1002 ---SEITDTVAASTAQSPGMN---LHSASLSLQQAFAFSELRRQAQMTGPNATPNFSHT 1055
Qy 1488 LETLPSTK---ESDSTOTTPSASCPSNSVNOVEDMEIETSEVKVKTSSPITSEESNL 1544
Dy 1056 GPTFPVVPVPLSLSLSTAGVPTTAATAAPVATSPSPND--ISTS-----VIQSEVTPTEGI 1109
Qy 1545 -----SNDFIDENGLPINKNENNVNGESKRKTIVTEVTMTSTVATESKTIVKEBGDKQT 1599
Dy 1110 AGVATSTGVVTSGLPI---PPVSESPVLLSVSSVSIITPAVVISITTSPLSQVSTSEI 1166
Qy 1600 VVSTENCAKSTVTTTTTIVTKLSTPSTGSGVDIISVKEQSKTVVTTVTDSLTTGGTL 1659
Dy 1167 VWSSTALYPSVTSATSA-----SAGGSTATPGPKPPA--VVSQQAAGS--TVVGATL 1215
Qy 1660 VTSMTVSKYSTDRDKVLMKFSRPPKTRSGTALPSKFKYTKSKTSKKSIFVLPNDLKKLA 1719
Dy 1216 TSVSTTTTSFFSTASQLSI-----QLSSSTSTSTPFLAETVVVSAH----- 1253
Qy 1720 RKGGIREVPVFNNAKPALDIWPYSPRPTFGITWRYRLQTVKSL--AGVSLMLRLLAS 1777

```
Db 1254 -----SLDKTSHS-----TTGLAFSLAPSSSSPGAGVSSYIS-----1288
QY 1778 LRWDMAAKVPPGGSTRTETSETTEITTEIKRDRDVGPIRFEYCIRKIIICPIGVPEPT 1837
Db 1289 -----QPGG-----LHPLVIPS 1301
QY 1838 PKETP-TPOKRLRSSALRPKRPETPKOTGPV-----LIETWAVE-----1877
Db 1302 IASTPILPOAAGFTPTPLLPQVPSIPPLVQPVANVPVQOOLTHSOPQOPALLPNOPHTHC 1361
QY 1878 -ELE-----LWEIRAFARVE-----KERAQAVEQAKRLEQO-----KPTV-----IATS 1918
Db 1362 PEVDSOTQPKAGIDDIKTLEKRLSLFSEHSSGAQHASVSLETSLVIESVTPTCIPPT 1421
QY 1919 TTSPT---SSTSTSTSPAKV-----WVAPISGVTTGKWLTVKVGSPATVTFQONKF 1971
Db 1422 AVAPSKLLTSTTCLPTNLPLGTVALFVTPVPTPG-----QVSTPVST-----1467
QY 1972 HQTFAWVKQGSNGVQVQKVLGIIPSTSTGTSQOFTSPQRTATVIRPNTSGSGG 2031
Db 1468 -----TSGVKPGTAPSKPPLTKAPLVPVGTLPAG 1497
QY 2032 TTSNSQVITGPIRPGMTVIRPLQO-----STLGRAIIRTPVMWOPG 2074
Db 1498 TLPSEQLPPP--GPSLTQSQPLEDLDAOLRRTLSPEXITVTSAVGPVMAAPTATEA 1555
QY 2075 A--POQVMTIIRQVSTAVSA-----PNTVSSTPGQS-----LTSAT 2112
Db 1556 GTOQKGVQVKGEPVLTATSSGAVFKMGFRQVVAADGAQKGNKSEDAKSVHFEST 1615
QY 2113 STSNIOSASOPRPOOGVKLTMAQLTQTOGHGNGOGLTVVIOGQOTTGOLQILPOG 2172
Db 1616 SESSVLSSS-----PESTLVKPE-----PNG 1637
QY 2173 VTLPQGGQOLMOAMPNTGVORFLETPLATATATATTT-----TTVSTTAAGTGE 2224
Db 1638 ITI-PGTTSDVPESAH-----KTTASEAKSDTGOPTKVGREFQVTTANKVGR 1683
QY 2225 QROSK-----LSPQMOVHODKTLPPAQSSSVGPAKAQOTAPQASRPOPOT 2270
Db 1684 FVSKTEKIDTKKGGPVASPPFMDLEQAVLP-----AVIPKKEPELSEFSLHNGPS 1738
QY 2271 PQO-----SPAQP-----2278
Db 1739 DPEAAFLSRVDGSGSPHSPHOLSSKSLPSQNLSQLSNSFNSYMSDNESDIEDL 1798
QY 2279 -----EVQ-----TQPEVQOTTVSSHYP-----SEAQPTHQS 2307
Db 1799 KLELRRLRDKHLKEIQDLSRQRKHEIESLYTKLGKVPVPAVIPPAPLSGRRRRPTKSKG 1858
QY 2308 SKPOVAA--QSOPO--SNVQGSFVRVQSPQSTRIRPSTPSOLSPOGQSOVQTTTSQPI 2362
Db 1859 SKSRSSSLGNKSPQLSGNLGSGSAASVLPQOT-LHP-----PG-----1897
QY 2363 PIQHTSLQIPSGQPO-SOPQVOS-STQTLSSGQTLN-QVSVSSRSPOLQIQP 2415
Db 1898 -----NIPESQNLQPLKPSPPSDNLYSAFTSDGAISVPSLSAPGQGIKOP 1945
```

```
RESULT 12
US-09-854-856-24
; Sequence 24, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; TITLE OF INVENTION: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
```

```
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2201)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-24

Query Match 2.5%; Score 374; DB 4; Length 2201;
Best Local Similarity 17.8%; Pred. No. 1.5e-13;
Matches 440; Conservative 302; Mismatches 764; Indels 968; Gaps 108;

QY 472 PDDPEQCKSEPTVEVGDKNVSANLGDNTNATSE-----ETSPSEGRSPVGCLESTP 526
Db 21 PAPAPKNGSSD--SSVGEK---LGAADAADAVTGRTEYRRRRHTMDKDSGAAATTTTTE 76
QY 527 -----DSS-----NMAEKKVASLPQDVPPEPNKTCESNTSATTSI- 564
Db 77 HRFFRRSVICDSNATALELPGLPLSPQSPAPVQSPAPPEPHR--EETVTATATSOVA 134
QY 565 -QPNLENSSSSELNSESASAKAADPEN--GERESHTPVYSIOEIVGDFTSEKSGEL 621
Db 135 QQP-----PAAAPGEQAVAGPAPSTVPSSTSKDRPVS-QPSLV-----GSK 175
QY 622 SESPGAGKAGSSTRITRLRNPDSKLSQKAAHAAHEANKLFKEGKEVLVVNSQGE 681
Db 176 EEPFPPARSQSGGS-----AKEPQBERSQ-----QQDD 203
QY 682 ISRLSTKKEVMKGNINNVFKLGQE---GKRVVHNOYSTNSEA-----LN 724
Db 204 IELETK--AVGMSNDGRFLKFDIEIGRSPKTVYKGLDTETTVEVANCELODRKLTKSE 261
QY 725 KQHREDHDKRHLAH-----KFCLTAPAGEFKWNGS-----VHGSKVLTII 764
Db 262 RQRFEKEAEMKGLQHPNIVREYDSWESTVKGKCVILVTELMTSGTLTKYLKRFKVMKI 321
QY 765 STLRLTITOLENNIPSSFFHPNW-----ASHRANW 794
Db 322 KYLRSWCROILKGL--QFLHTRTPPIIHRDLKCDNIFITGPTGVSVKIGDLGLATLKRASF 379
QY 795 IKAV-----QMSCKPRE-----FALALAILECA-----V 818
Db 380 AKSVIGTPEFMAPEMVEEKYDESVDYAFGCMCMLEMATSEYPYSECQNAAYIRRVTSQV 439
QY 819 KPV-----VMLPIWRE-----FLGHTRLHMTSI-----EREKEK-- 849
Db 440 KPASFQKVAIPEVKETIEGCIQONKDERYSIKDLNHAFFOEBTGVRVELAEDDEGKIA 499
QY 850 -----VKKKEKKOEETEMQOATVVKYFPFKHVKQKQGEYRVTYG---GGWSW 897
Db 500 IKLWLRIEDIKLLKGYKNEA-----IEFSFDLERDVPEDVAQEMVSGVCGDHKT 553
QY 898 ISKTHYRVFVKPLPGNTNVNYRKSLEGTNNNDENNDKSKRSRPAKKIKIEPDSEKD 957
Db 554 MAKI-----IKDRVSLIKRKRQQLVREDEK-----KOEESLQK 591
QY 958 EVKGSAAKAGADQENMDISKITEKKDDQVKELLDSDSKPCKPEPMEVDDDMKTESHVNC 1017
Db 592 QVEQSSASQ-----TGIKQLPSASTGIP-----TASTTSA 621
QY 1018 QESSQVDVNVNVEGPHLRTSYKKKTKSKLDGLLE-----RRIKQFTLEEKQL 1066
Db 622 SVSTQVEPEPEADQHQOQYQOQPSISVLSDGTVDSGQSSVFTESRSVQQTQVSYGSH 681
QY 1067 ERTKLEGGIKG-----IGKTSNSSKN 1088
```


QY 508 EETSPGSRPYGCLSETPDSSNMAEKKVASELPQDVPPEPNKTCESSTNTSATTTSIOPN 567
D 508 EETSPGSRPYGCLSETPDSSNMAEKKVASELPQDVPPEPNKTCESSTNTSATTTSIOPN 567
D 62 EETVATATSOVA--QOPPAAPAGEQAVAGAPSTVP-----SSTKDRPVSPQS 110
QY 568 LENSNSSELNQSOSAKAADPENGERSHTPVSIQEEIVG-----DFTSEKSTG 619
D 568 LENSNSSELNQSOSAKAADPENGERSHTPVSIQEEIVG-----DFTSEKSTG 619
D 111 LVGSKPEPPARSGGGGSAKEPOERSOQODDIELETKAVGMSNDGRFLKFDIEIGR 170
QY 620 ELSSEPGAGGASGSTR1-ITRLRNPDSKLSQKSOVAAAHEANKLFKEKEKVELVYNS 678
D 620 ELSSEPGAGGASGSTR1-ITRLRNPDSKLSQKSOVAAAHEANKLFKEKEKVELVYNS 678
D 171 SFKT-----VYKGLDTETFEVANCELODRKLTATSEKOR-----FKBAEMLKGLQ 216
QY 679 QG8ISRLSTKKEVIMKGN--INNYFKLQEGRYRVHNOYSTNSFALNKHORED----- 731
D 679 QG8ISRLSTKKEVIMKGN--INNYFKLQEGRYRVHNOYSTNSFALNKHORED----- 731
D 217 HPNIVFYDSWSTVYKGGKCIVLVTGLTSGTLKYLKFKYKIMKVLKRSWCRLKGLQ 276
QY 732 --HDKRRHLAHK-----FCITPAGEFKWNGSVHSGKVLTTSTLRLTTQLENNIPSSF 782
D 732 --HDKRRHLAHK-----FCITPAGEFKWNGSVHSGKVLTTSTLRLTTQLENNIPSSF 782
D 277 FLHTRTPPIIHRDLKCDNIFITGPTGTVKIGD-----LGLATLK----- 315
QY 783 FHPNASHRANKTAV-----QMCCKPRE-----FALALAILLECA----- 817
D 783 FHPNASHRANKTAV-----QMCCKPRE-----FALALAILLECA----- 817
D 316 -----RASAKSVIGTPEFMAPEMEYEEKYDESVDVYAFGCMLEMATSEYPTSECON 367
QY 818 -----VKPV-----VMLPIWRE-----FLGHTRLRHMTSI-- 842
D 818 -----VKPV-----VMLPIWRE-----FLGHTRLRHMTSI-- 842
D 368 AAQIYRRVTSVYKPAFDKVAIPEVKEIIEGCIQONKDERYSIKDLNHAFFQEETGVRV 427
QY 843 ---EREKEK-----VKKKEKQEBEETMOQATWVYTFPVKQVQWVKQKEEYR 888
D 843 ---EREKEK-----VKKKEKQEBEETMOQATWVYTFPVKQVQWVKQKEEYR 888
D 428 ELAEEDGGEKIAIKLRIEDIKKLGKYKDNEA-----IEFSDLRDVEDVVAQEMV 481
QY 889 VTGY---GWSWTSKTHVRYFVPLPCNTNVNRYKSLGKTKNMNDMDNDESKRKCGRP 945
D 889 VTGY---GWSWTSKTHVRYFVPLPCNTNVNRYKSLGKTKNMNDMDNDESKRKCGRP 945
D 482 ESGYCEGDHKTWAKA-----IKDRVSLIKRKRQOLVREEQEK----- 522
QY 946 KTIKIEPDSEKDEVKGSDAAKADQNMEDISKITEKKQDQVKEKLLSDSDKCKEPPMEV 1005
D 946 KTIKIEPDSEKDEVKGSDAAKADQNMEDISKITEKKQDQVKEKLLSDSDKCKEPPMEV 1005
D 523 ---KQESSLIKQOQESSASQ-----TGIKOLPASTGIP----- 554
QY 1006 DDMKTEHVNCOFESSQDVVNVSEGHRTSYKKKTKSKLDGLLERIKQFTLEKOR 1065
D 1006 DDMKTEHVNCOFESSQDVVNVSEGHRTSYKKKTKSKLDGLLERIKQFTLEKOR 1065
D 555 -----TASTTSASVTOVEPEEPADQHOLOLQOQPSISVLSGTVD----- 596
QY 1066 LEKIKLEGGIKGIGKSTNSKNLSESPVITKAKEGCSQD-----SMRQEOS 1112
D 1066 LEKIKLEGGIKGIGKSTNSKNLSESPVITKAKEGCSQD-----SMRQEOS 1112
D 597 -----SGOGSSVFETESRVSSQTV-----SYGSQHEQAHSHTGVPHIPSTVQAOS 642
QY 1113 PNANDOPEDLIQCCSQSDSSVLRLMSDPSHTTNKLYPKDRVLDVDSIRSPETKCPKQNSI 1172
D 1113 PNANDOPEDLIQCCSQSDSSVLRLMSDPSHTTNKLYPKDRVLDVDSIRSPETKCPKQNSI 1172
D 643 QPHGVYPPSSVAQSQGOGPSSSSSLGTVSSSQPIQHPOQ-----OGIQQTAPPQQT 695
QY 1173 ENDI-----EEKVSDLASRGOEPT-----KSKTKGNDFIDDSKLASADDIGTLI 1217
D 1173 ENDI-----EEKVSDLASRGOEPT-----KSKTKGNDFIDDSKLASADDIGTLI 1217
D 696 QYLSQTSSTSEATTAQVVSQAPQVLPQVVSAGKOSTQG-----VSQVAPAEVAV-- 747
QY 1218 CNKKKPLIQEESDTIYSSSKSALHSSVPKSTNDRDATPLSRAMDFEGK----- 1265
D 1218 CNKKKPLIQEESDTIYSSSKSALHSSVPKSTNDRDATPLSRAMDFEGK----- 1265
D 748 ---AQP---QATQPTTLASSVDSAHSDVAGMSDGNENVPSSSGRHEGRTTKRHYKSVRS 802
QY 1266 -----LCDSSESNTLENSDT-----VSTQ-----DSSEE-----DM 1293
D 1266 -----LCDSSESNTLENSDT-----VSTQ-----DSSEE-----DM 1293
D 803 RSRHEKTSRPLKRLVNSKNGDRVVECOLETHNRKVMVTFKFDLDGDNPEEATIMVNNDF 862
QY 1294 IVQNSNESISEQRTREODVEVLEPLKCELVSGESTGNCBDRPLPVKGTANGKKPSOOKK 1353
D 1294 IVQNSNESISEQRTREODVEVLEPLKCELVSGESTGNCBDRPLPVKGTANGKKPSOOKK 1353
D 863 ILAIERESFVDQVR---EIEKEADEMLSEDSVSEPEGD---QGLSELOGDDYGEFGSQKLE 918
QY 1354 LEER---PVNKCSDQIKLKNNTDKNNENRESEKKGQRTSTFOINGKDNKPKYILKGE-- 1408
D 1354 LEER---PVNKCSDQIKLKNNTDKNNENRESEKKGQRTSTFOINGKDNKPKYILKGE-- 1408
D 919 GEFKQPIASSMPQOIGIPT-----SSLTQV-----VHSAGRFR 952

QY 1409 CLKEISESRVSGNVEPKVNNINKIIPENDIKSLJVKESAIRPFINGDVIMEDFERNSS 1468
D 953 IVSPVPESLRESKVEP-----SEITDTVAASTAQSPGMN---LSHSASSLSLQ 998
QY 1469 ETKSHLLSSDAEG---NYRDSLETLPSTK---ESDSTQTTWPSASCSPESNVQVE 1519
D 999 QAFSELRAQMTPEGNTAPPNFSHTGPTFFVVPFLSSTAGVPTTAATATAPVATSPPN 1058
QY 1520 DMEIETSVKVKYTSPTSEEENL-----SNDFIDENGLPINKNNVENGESKRTVITE 1574
D 1059 D--ISTS---VIOSEVIVPTEEGTAGVATSGVVTSGGLPI---PPVSESPLSSVSVSS 1109
QY 1575 VTTMTSTVATESKTVIKVEKQKQVVSSTENCASKSTVTTTNTTKLSTPSTGSGVDII 1634
D 1110 ITIPAVVSISSSTPSLOVPTSTSEIVVSSSTALYPSVTSATSA-----SAGGSTATP 1161
QY 1635 SVKEOSKTVVTTVTDLSLTGTLVTSMTVSKYSTKRDVKVLMKFSRPPKRSSTALPS 1694
D 1162 GPKPPA--VVSQAAAGS--FTVGATLTSVSTTTSTFFSTASQLSI-----QLSSSTSTPT 1211
QY 1695 YRKFTVKSTKKSIFVLPNDLKLARKGGIREVPYFYNNAKPALEDIWPYPSPRPTFGITW 1754
D 1212 LAETVVVSAH-----SLDKTSHSS---TTGLAF 1236
QY 1755 RYRLQTVKSL--AGVSLMLRLLWASLRDMDMAKVPPGGSGTRETSETTEITTEILKRR 1812
D 1237 SLSAPSSSSSPGAGVSSYIS-----QPGG----- 1260
QY 1813 DVGYPYRIRYCIKRIKIPGVETPKETP--TPQRKGLRSSALRPPKRPETPKOTGPV--- 1868
D 1261 -----LHPLVIPSIVASTPILPQAAGPTSTPILLQVPSIPLVQPVANV 1304
QY 1869 --IIEETWAE-----ELE-----LWEIRAFARVE---KKAQAV 1899
D 1305 PAVQOTLIHSQOPALLPNQPTHCPEDVSDTPQKAPGIDDIKLEKRLSFSEHSSG 1364
QY 1900 EQAKKRLSEQ---KPTV---IATSTTST---SSTTSTISPAKV---MWAPITSGSVT 1946
D 1365 AQHASVLSLETSLVETSTPGITTAAPASKLLSTTCLPPTNPLGTVALPVTPVVT 1424
QY 1947 TGTGMVLTTKVGSAPATVTFQONKNFHQTFATVWVQKQSGVGVVQVQKVLGIIPSTGTS 2006
D 1425 PG-----QVSTPVSTT-----TSGVK 1440
QY 2007 QOTFTSFQPRATVTRPNTSGSGTSSNSQVITGPQIRPGMTVIRTPLOQ----- 2057
D 1441 PGTAQSKPLTKAPVLPVGTLPAGTLPSEQLPPFP--GPSLTQSOQPLEDLDAQLRRTL 1498
QY 2058 -----STLKGKAIIRTPVMVOPGA--POQVMTQIIRGQPVSTAVSA-----PNTV 2099
D 1499 SPEXITVTSAGVPVSMMAAPTATEACTQPKQVQKGVQKEGVLATSSGAGVFKMGRQVSV 1558
QY 2100 SSTPGQKS-----LTSATSTSNIOSSASQPPRPOQGVKLTMAQLTQLTQGHG 2147
D 1559 AADGAQKEGKNKSEDAKSVHFESESTSESSVSSSS---PESTLVKPE----- 1602
QY 2148 GNOGLTWIQQOGQTTQLOLIPQGVTVLPFGCQQLMQAMPNGTVQRFELTFLATTATT 2207
D 1603 -----PNGITI--PGISSDVPESAH-----KTTASE 1626
QY 2208 ASPTT-----TTVSTTAAGTGEQROSK-----LSPQMOMVHQDKTLP 2245
D 1627 AKSDTCQPKVGRFQVTTTANKVGRFSVSKTEKIDTKKEGVPASFPFMDLEQAVLP-- 1684
QY 2246 QSSSVGPAKAQPOTAQPSARPOQTOPO-----SPAQP----- 2278
D 1685 ---AVIPKKEKPELSEPSHLNGSPDPEAAFLSRDVGSGSPHQLSSKSLPSONLS 1741
QY 2279 -----EVO-----TOPEVQQTIVSSH 2295
D 1742 QSLNSFNSSYMSDDNESDIEDLKLRLRRLKHLKEIQDLOSKQKHEIESLYTKLCK 1801
D 2296 VP-----SEAQPTHQASSKQPVAA---OSQPO---SNVQOGSPVRVQSPQSTR 2337

Db 1802 VPAVPIPPAALSGRRRTTKSGKSSRSSLGNKSPOLSLGQSAASVHLHPQOT- 1860
Qy 2338 IRFSTPSQLSPGOQSOVQTTSOPIPIQHTSLQIPSOQPO- SQPOVQS- STQTLSSGQ 2395
Db 1861 LHP-----PG-----NIPESGQQLLOPLKPSFSSDNLYSAF 1892
Qy 2396 TLN-QVSVSPSRPQLOIQOP 2415
Db 1893 TSDGAISVPSLSAPGQGIKOP 1913

RESULT 14
US-09-854-856-62
; Sequence 62, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 2048
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2048)
; OTHER INFORMATION: xaa = Any Amino Acid
US-09-854-856-62

Query Match 2.5%; Score 370; DB 4; Length 2048;
Best Local Similarity 18.2%; Pred. No. 2.3e-13;
Matches 419; Conservative 287; Mismatches 746; Indels 846; Gaps 96;

Qy 508 EETSPSEGRSPVCLSETPDSSNMAEKKVASELPQDVPEPKNKTCBSSNTSATTTISQPN 567
Db 62 EETVTATATSOVA--QQPPAAAPAGEQAVAGPAPSTVP-----SSTSKDRPVSQPS 110
Qy 568 LENSNSSELNNSQSESAKAADDPNGERESHPTVSIQEEIVG-----DFTSEKSTG 619
Db 111 LVGSKPEPPPARSGSGGSAKEPQERSQQDDIELETAKVAGMSNDGRFLKPDIEITGRG 170
Qy 620 ELSSEPCAGKAGSGSTRI-ITRLRNPDLSKLSQSOVAAAHAANKLPKEGKEVLVNS 678
Db 171 SFT---VYGLDTETVEVWCELQDRKLITKSERQ-----FKEEAEMLKGLQ 216
Qy 679 QGEISRLSTKKEYIMKGN--INNYFKLGOEGKRVYHNOYSTNSFALNKHQHRED----- 731
Db 217 HPNIVRFYDSEWSTVKGKCIVLVTELTSGTLTKYLRKFKVMKIKVLRSCWQILKGLQ 276
Qy 732 --HDKRRHLAHK-----FCLTPAGEFKWNGSVHGSKVLTITSLRLTITQLENNIPSSF 782
Db 277 FLHTRTPPIIHRDLKCDNIFITGPTGSVKIGD-----LGLATLK----- 315
Qy 783 FHPNASHRANWIKAV-----OMCSKPRE-----PALALALECA----- 817
Db 316 -----RASFAKSVIGTFEPFMAPMEYEEKYDESVDVYAFGCMLEMATSEVPYSECON 367
Qy 818 -----VKPV-----VMLPIWRE-----FLGHTRLHRMTSI-- 842
Db 368 AAQIYRRVTSYGVKPSAFDKVAIPEVKEIIEGCIQONKDERYSIKDLLNHAFFOETGVRV 427
Qy 843 ----EREKEK-----VKKKEKQEEETMQOATWVYTFPVKQHVWKQGEYR 888

Db 428 ELAEEDGGERIAIKLWLRIEDIKLKGKYDNEA-----IEFSFDERDVPDVAQEMV 481
Qy 889 VTGV---GMSWTSKTHVYFVPLPGTNTVNYNFKSLEGTKNNNDENNDSDKRCRSRP 945
Db 482 ESGVVCEDDKTWAKA-----IKRVSILKRRQRLVREDEKK----- 522
Qy 946 KKIIEPDSKDEYKSDAAKAGADONEMDISKITEKKDQDVKELLDSDKPKCKEPMEV 1005
Db 523 ---KQESSLKQVQEQSSASQ-----TGIKQLPSASTGIP----- 554
Qy 1006 DDMKTESHVNCQESSQDVVNVSEGHFLRTSYTKKTKSKLDCGLLERRIKQFTLEBKQR 1065
Db 555 -----TASTTSASVSTQVEPEPEADHQQLQYQOPSIVLSDCGTVD----- 596
Qy 1066 LEKIKLEGGIKGKTSTNSSKNLSESPVITKAKEGCQSDSMRQEQSPNANDOPEDLIQ 1125
Db 597 -----SGQSSVFTESRV-----SSQTVSYGSGOHAHSTGT 629
Qy 1126 GCSQSDSVLRMSDPSTHTNKLYPKDRVLDDVSIRSPTKCPKQNSIENDI-----EE 1178
Db 630 VPGHIPSTVQAQSQP-----HGVYPPSSVQOGI-----QQTAPPOQTVOYLSQTSSTSEA 680
Qy 1179 KVSOLASRGQEP-----KSKTKGDNDFIDDSKLASADDITGLCKKKNKPLIQESD 1230
Db 681 TTAQPVSQPAQVLPQVSAGKOSTQG-----VSQVAPAEPAVAV-----AQP--QATQP 727
Qy 1231 TVSSSKSALHSSVPKSTNDRDATPLSRAMDFEKG----- 1265
Db 728 TTLASSVDSAHSDVASGMSDGNENVPSSGRHGRITKRYRKSVRSRHKEKTSRPKLR 787
Qy 1266 -LGCDSNSTLENSDT-----VSIO-----DSSEE-----DMIVONSNESISEQF 1306
Db 788 ILNVNKGDRVVECOLETHNRKMVTFKFDLDGDNPEIATIMVNDIFILAIERESFVDQV 847
Qy 1307 RTREQDVEVLEPLKCELVSGESTGNCEDRLPVKGTENGAKKPSQOKKLEIR---PVNKCS 1363
Db 848 R---EIEKADEMLSEDVSVPEEGD-QGLESLOKDDYGFSGSKLEGEFKQPIPASSMP 903
Qy 1364 DQIKLKNWTDKNNENRESEKKGORTSTFOINGKDNKPKYLKGE--CLKEISRSRVVSG 1421
Db 904 QQIGIPT-----SSLTQV-----VHSAGRRRIVSPVPSRURES 937
Qy 1422 NVBPKVNNINKIIPENDIKSLTVKESAIRPFINGDIVIMEDFNERNSETSKHLLSSDAE 1481
Db 938 KVP-----SEITDTVAASTAQSPGMN---LSHSASSLSLQQASELRAQWTE 983
Qy 1482 G-----NYRDSLETLPSTK---ESDSTQTTTSPASCPESNSVNOQVDEMEIETSEVKVT 1532
Db 984 GPNTAPPNFSGHTGTFPVVPPFLSSAGVPTTAAATAPVATSPSPND--ISTV-----VI 1037
Qy 1533 SSPITSEESNL-----SNDFIDENGILPKNNENNGESKRKVTITVETMTSTVATESK 1587
Db 1038 QSEVTPTEGIAAGVATSGVTTGGGLPI---PPVSESPVLSVSVSITIPAVVSIISITTS 1094
Qy 1588 TVIKVEKGDQTVVSVSTENCACAKSTVTTTTTTLSTPSTGSGVDIISVKEQSKTVVTTT 1647
Db 1095 PSLOQVPTSTSEIVSVSTALYPSVTSATSA-----SAGGSTATPGKPPA--VVSOQ 1144
Qy 1648 VTOSLTTTGTGLVTSMTVSKYSTRDKVKLMKFSRPKKTSGTALPSYKRFVTKSTKSI 1707
Db 1145 AAGS--TTVGATLTSVSTTTFSPSTASOLS-----QLSSSTSTPTLAETVVVSAH--- 1193
Qy 1708 FVLPNDDLKLRKGGIREVPYFNNAKPAIDWIPYSPRPTGITWRYRLQTVKSL--A 1765
Db 1194 -----SLDTSKSHS-----TTGLAEFSLAPSSSSSPGA 1221
Qy 1766 GVSMLRLLWASLRWDDMAKVPVGGSGSTRTETSETTEITTEIKRRDVGVPYIRFEYCI 1825
Db 1222 GVSSYIS-----QPGG----- 1232
Qy 1826 RKIICIGVPEPKETP--TPQRKGLRSSALRPKRPTPKOTGPV-----IETWVAEE-- 1877

Db 1233 ---LHPLVPSVIASTPILPOAAGTSTPLLPQVPSIPPLVQPVANVPAVQOITLHSQPQ 1289
QY 1878 -----LWEIRAFARVE---KEQAQVQQAQKLEQO-- 1910
Db 1290 PALLPNQPHHCPEVDSQTQKACIDIDIKLEKRLSFLSEHSSGGAQHASVLETSLV 1349
QY 1911 -KPTV---IATSTTSPT---SSTTSTISPAQV---MVAPISGVTITGKRWLTTTKVGS 1959
Db 1350 IESVTPGIPITAVAPSKLLSTSTCLPPTNLPLGTVALPVPVPTPG-----QVST 1402
QY 1960 PATVTFQONKFNHQTATWVQKQSGNSGVVQVQKVLGIIPSSGTSTQOITTSQPTAT 2019
Db 1403 PVSTT-----TSQVKGPTAPSKPPLTKA 1425
QY 2020 VTIRPTSGGTTNSQVITGPQIRPGMTVIRTPLOQ-----STLGK 2062
Db 1426 PVLPGVTELPAGTLPSEQLPFPF--GPSLTQSOQPLEDLDAQLRRLTSPXITVTSAGP 1483
QY 2063 AIIRTPVMVQPCA--PQVMTQIIRGQPVSTAVSA-----PNTVSSTPGQKS----- 2107
Db 1484 VMAAPTATEAGTQPKGVQVKGEPVLATSSGAGVFKMGRFQVSVAAQAKGEGKNS 1543
QY 2108 -----LTSATSNIOSSASQPRPOQGVKLTMAQLTOLTQCHGNOGLTVVIOQOG 2160
Db 1544 EDAKSVHFESTSESSVLSSS-----PESTLVKPE----- 1574
QY 2161 QTTGQLLIPQGVTVLPQPGQOLMOAAMPNGTVQRFLEFTPLATTATTATTT----- 2212
Db 1575 -----PNGIIV-FOISSDVESA-----KTASEAKSDTGQTKVGR 1611
QY 2213 TTVSTTAAGTGEQROSK-----LSPOMOVHQDKTLPPAGSSVGVPAKAQOQ 2258
Db 1612 FOVTTTANKVGRFSVSKTEDKITDTKKEGVPVSPFMDLEQAVLP-----AVIPKKEKE 1666
QY 2259 TAQPSARQPOQPO-----SPAQP----- 2278
Db 1667 LSEPHLNGSPDPEAFLESRDVGSGSPHSPHOLSKSLPSQNLSSNSFNSYMS 1726
QY 2279 -----EVO-----TOPEVQOTQTVSSHP----- 2297
Db 1727 SDNESDIEDLKLRLRLDKHLKEIQDLSQKHEITESLYTKLGKVPVAVIIPPAAPL 1786
QY 2298 --SEAQPHAGSSKQVAA---QSQPO---SNVQSGSPVRVQSPQSTRIRPSTPSQLSPGQ 2350
Db 1787 SGRRRRTPKSGKSSRSSLGNKSPQLSGNLSGQSAASVHLHPQOT-LHP--PGNIPBSG 1843
QY 2351 OSOVQTTTSQPIPIOPHT-----SLQIPSGOGOPQOPQVQSTOTLSSQOTLNO 2399
Db 1844 QNQL-----LQPLKPSFSSDNLVSFTSDGALSVPLSLAP---GQGTSTNTV--GATVNS 1894
QY 2400 VSVSSPSRPLQIQOQPQ 2417
Db 1895 -----QAQAQAP 1901

RESULT 15
US-09-854-856-8
; Sequence 8, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8
; LENGTH: 2229
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2229)
; OTHER INFORMATION: xaa = Any Amino Acid
US-09-854-856-8

Query Match 2.5%; Score 369; DB 4; Length 2229;
Best Local Similarity 17.9%; Pred.No. 3e-13;
Matches 443; Conservative 306; Mismatches 781; Indels 948; Gaps 111;

QY 472 PDDPEQCKSEBPEVGVGDKGNSVANLGDNTTNAISE-----ETSPSEGRSPVPGCLSTP 526
Db 21 PAPAPKNGSSD-SVGEK---LGAAAADAVTGRTEEYRRRHMTMDKSRGAAATTTTTE 76
QY 527 -----DSS-----NMAEKKVASELPQDVPEEPNKTCCSSNTSATTTTISI- 564
Db 77 HRFFRRSVICDSNATALEPLGLPLSLPQPSIPAAPVQSAPEPHR--EETVTATATSQVA 134
QY 565 -QPNLENSSSSELNSSOSASAKAADDPEN--GERESHTPVYSIOEIVGDFTSEKSTGEL 621
Db 135 QQP-----PAAAQPEQAVAGPAPSTVPSSSTSKDRPVS-QPSLV-----GSK 175
QY 622 SESPGAGKAGSGSTRITRLRNPDSKLSQKSSQVAAAAHEANKLFKEGKVLVYNSQGE 681
Db 176 EEPFARSQSGGS-----AKEQBERSQ-----QQDD 203
QY 682 ISRLSTKKEVIMKGNINNYFKLQBP---GKYRVYHNYSTNSFA-----LN 724
Db 204 IEELETK--AVGMSNDGRFLKFDIEIGRGSFKTVYKGLDTETTVEVAVCELQDRKLTKSE 261
QY 725 KHOHREDHDKRRHLAH-----KFCLTTPAGEFKWNGS-----VHGSKVLTI 764
Db 262 RQRFKEEAEMKLGLOHPNIVRFYDSWESTVVGKKCIVLVTTELMTSGTLTKYLKRFKVMKI 321
QY 765 STLRLTITQLENNIPSSFFHPNW-----ASHRANW 794
Db 322 KVLRSWCQILKGL--QFLHTRTPIIHRDLKCDNIFITGPTGYSVKIGDLGLATLKRASF 379
QY 795 IKAV-----QMSKPRE-----FALALAILCA-----V 818
Db 380 AKSVIGTPEFMAPEMYEKEYDESVDVYAFGCMCMLEMATSEYPYSECQNAAYIRVTVSGV 439
QY 819 KPV-----VMLPIWRE-----FLGHLRLHMTSI-----EREKEK- 849
Db 440 KPASFDKVAIEVKELIEGCIQNKDERYSIKDLLNHAFQOETGVRVLAEEDEGKIA 499
QY 850 -----VKKKEKKQEEETMQOATWVYTFPVKQVQKQKQGEYRVYGY---GGWSW 897
Db 500 IKLWLRIEDIKLKCKYKDNEA-----IEFSFDLERDVPEDVAQEMVESGVCEGDHKT 553
QY 898 ISKTHVYRFVKLPNTVNVYRKSLGCTKNNDENM--DESDKRRKCSRPKKIKIEPDSE- 955
Db 554 MAK-----IKDRVSLIKRKRQORQVREEQKKQOESLSKQOQVQSSAS 599
QY 956 ----KDEVKGSDAAGADQNMEDISKITEKKDQDKELLDSDSKPCKEEPMVEVDDD--- 1008
Db 600 QTGIKQLPSASTGIFTASTTSASVSTQVEPEPEA----DQHQOQLQYQPSISVLSDGTV 655
QY 1009 -----MKTESHVNCQES-----SOVDVVNVYSEGHRTSYKK 1040
Db 656 DSGQGSQSVFTESRVSSQQTVSYSQHEQAHSHTGTVPGHIPSTVQASQPHGVYPPSSVAQ 715
QY 1041 -----KTKSSKLDGLLERRIKOFTLEEKORLEKIKLEGGIK-----GICKTSTN 1084
Db 716 GQSQGQPSSSSLTG-----VSSQPIOHQOQOQOQIQTAPQOTVQVLSLSQTSIS 765
QY 1085 SSKNLSSES-----PVITKAKGQCSQSMRQEQSPNANNNDPEDLIQ--CQSDSS 1133
Db 766 SEATTAQVPSQPAQVLPQVYSAGKQG--FPPRLPQYPGDSNIAPSNVAVSVCIH--ST 821

QY 1134 VLKMSDPST-----TNKLYPKDRVLDVDSVSTRSBETKCPKONSTENDIE 1177
Db 822 VLXPPMPTEVLATPGFYFTVQPYVESNLLVPMGGVGQVQVSPG- 867
QY 1178 EKVSDLASRGQPTKS-----KTKGNDDFFIDDSKSLASADDIGTLICKNKKPLIOEESD 1230
Db 868 -----GSLAQAPTSSQAVLESTQG-----VSQVAPAEPAV-----AQP--QATOP 908
QY 1231 TIVSSKSAHSVPKSTNDRDATPLSRAMDFEGK----- 1265
Db 909 TULASSVDSAHSDVAGMSDGNENVPSSGRHEGRTKHKYKSVRSRSHKTSRPKLR 968
QY 1266 -LGCUSENSTLENSDT-----VSIQ-----DSSEE-----DMIVQNSNESISEQF 1306
Db 969 ILNVNKGDRVVVECOLETHNRRKMTVKFDLDGDNPEIATIMVNDFILAIRESFVDOV 1028
QY 1307 RTREODVEVLEPKCLVELYSGESTGNCEDRLPVKGTGANGKPKSQOKKLEER--PVNKCS 1363
Db 1029 R---ELIEKADEMLSEDVSVEPEGD--OGLESQGGDDYGFSGQKLEGEFKOPIPASSMP 1084
QY 1364 DOIKLKNTDKKNNRESEKKGORTSTFQINGKDNKPKIYLKGE--CLKEISESRVWSG 1421
Db 1085 QOIGIPT-----SSLTQV-----VHSAGRRFIVSPVPESRLRES 1118
QY 1422 NYEPKVNINKIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSSETKSHLLSSSDAE 1481
Db 1119 KVFP-----SEITDTVAASTAQSPGMN--LSHSASSLSLQQAFAFSELRAQMT 1164
QY 1482 G-----NYRDSLETLPSK-----ESDSTQTTTPSASCPSNVQVEDMEIETSEVKVT 1532
Db 1165 GNTAPPNFSHTGPTFVVPPLSSIAGVPTAAATAPVPATSSPND--ISTS-----VI 1218
QY 1533 SSPITSEESNL-----SNDFIDENGLPINKNNVNGESKRKTIVITEVTMTSTVATESK 1587
Db 1219 QSEVTPTEEGTAGVATSGVVTSGGLPI---PPVSESPVLSVVSSITIPAVVISITS 1275
QY 1588 TVIKVEKDKQVWVSTENCASKSTVTTTTTKLSTPSTGSGVDIISVKESKTVVTTT 1647
Db 1276 PSLOVPTSTSEIWSSTALYPSVTVSATSA-----SAGGSTATPGPKPPA--VVSQ 1325
QY 1648 VTDSLTTGGTLVTSMTYSKEYSTRDKVKLMKFSRPPKTRSGTALPSYRKFVTKTKKSI 1707
Db 1326 AGS--TTVGATLTSVTTSTFFSTASQLSI-----QLSSSTSTPTLAEATVVVSAH--- 1374
QY 1708 FVLNDDLKARKGGIREVPYFNTYNAKPALDIWPPSPRPTFGITWRYRLQTVKSL--A 1765
Db 1375 -----SLDKTSHS--TTGLAFSLSAPSSSSPCA 1402
QY 1766 GVSMLRLWASLRWDDMAKVPPGGGSTRTETSETTEITTEIIKRRDVGPGYIRFEYCI 1825
Db 1403 GVSSYIS-----QPGG----- 1413
QY 1826 RKIIICPIGVPTPKETP--TPQRKGLRSSALRPKRPETPKQTGPV-----IETWVAEE-- 1877
Db 1414 ---LHPLVIPSIVASTPILPOAGTSTPLPQVPSIPPLVQPVANVPAVQOTLIHSQP 1470
QY 1878 -----ELE-----LWEIRAPAEERVE---KEKAQAEQQAOKKLEEQ-- 1910
Db 1471 PALLPNQPHTHCPEVDSOTQKAPGIDDIKTLEEKLSLFSHSSSGAQAHSVLETSLV 1530
QY 1911 -KPTV---IATSTTSPT---SSTTSTISPAQKV-----MVAPISGSVTTGTKMYLTTKVGS 1959
Db 1531 IESTVTPGTPTAVAPSILLTSTTCLPPTNLPLGTVALPVTVPVTPG-----QVST 1583
QY 1960 PATVTFQONKNFHQTFATVWVGQSQNSGVVQVQKVLGIIPSTGTSQQTFTSFQPTAT 2019
Db 1584 PVSTT-----TSGVKPCTAPSKPPLTKA 1606
QY 2020 VTIRPNTSGSGTTSNSQVITGPQIRPGMTVIRTPLQ-----STLCK 2062
Db 1607 PVLPGVTELPACTLPEQLPPFP--GPSLTQSQQPLEDLDAQLRRTLSPEXITVISA VGP 1664

QY 2063 AIIRTPVMVOPGA--PQVMTQIIRQPVSTAVSA-----PNTVSSTPGOKS----- 2107
Db 1665 VSWAAPTAITEAGTQPKGVSOVKEGFLVATSSGAGVFRMGRFQVSVAAADGAQKEGKNKS 1724
QY 2108 -----LTSATSTSNIOSSASQPPRQOGOVKLTMAQLTQLTQGHGGNGLTVVIQGG 2160
Db 1725 EDKSVHFESSSESLSVSSS-----PESTLVKPE----- 1755
QY 2161 OTTGQIQLIPQGVTVLPGFQOOLQMAAMPNGTVQVRELFPLATATTASTTT----- 2212
Db 1756 -----PNGITI--PGISSDVPESAH-----KITTAESAKSDTQPTKVGR 1792
QY 2213 TTVSTTAAGTGEORQSK-----LSPQMVOHQDKTLPPAQSSSVGCPAKAQPO 2258
Db 1793 FQVTTTANKVGRFSVSKTEDKITDTKKEGFVASPPFMDLEQAVLP-----AVIPKKEKPE 1847
QY 2259 TAQPSARPOPOQPO-----SPAQP----- 2278
Db 1848 LSEPSHLNGPSSDPPEAAFLSRDVGSGSPHSPHQLSKSLPSONLSQSLNSFNSSYMS 1907
QY 2279 -----EVQ-----TOPEVQTQTTVSSHVP----- 2297
Db 1908 SDNESDIEDLKLRLRLRDLKHLKEIQDLQSRQKHEIESLYTKLKGVPVPAVIIPPAAPL 1967
QY 2298 ---SEAQPTHAQSSKPOVAA---OSQPQ---SNVOGQSPVRVQSPQTRIRPSTPSQLSPGQ 2350
Db 1968 SGRRRRPTKSGKSSRSRSLGNKSPQLSGNLGQSAASVYLHPQOT--LHP--PGNIPESG 2024
QY 2351 OSQOVQTTTQPIPIOPHT-----SLOIPSOQOQSOPOQVOSSTQTLSSGOTLQ 2399
Db 2025 QNQL-----LQPLKPSONLYSAFTSDGAISVPSLSAP---GQGTSTNTV--GATVNS 2075
QY 2400 VSVSPSRPQLQIQOPQ 2417
Db 2076 -----QAAQAQ 2082

Search completed: September 24, 2003, 01:18:58
Job time : 53.9984 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 00:30:12 ; Search time 69.5063 Seconds
(without alignments)
10792.671 Million cell updates/sec

Title: US-09-698-295-1

Perfect score: 14971

```
Sequence: 1 MVSEEEEEEDGDAEETQDSE.....KLKGFKASRSHNNKLQSTAS 2907
```

Scoring table:

Gapop 10.0 , Gapext 0.5

0
:
0
0
4
0
2
5
0

-
0
:
0
4
2
0
5
0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Minimum DB seq	length: 0
Maximum DB seq	length: 2000000000

Post-processing: Minimum Match 0.8

Post-processing: Minimum Match 0%
Maximum Match 100%

Maximum Match 100%

Listing first 15 summaries

Database :

```

1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.ornamentele:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.yrvis:*
16: sp.bacteriap:*
17: sp.archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %			ID	Description
	Score	Match	Length		
1	14170	94.6	2781	4 Q9UIG2	Q9UIG2 homo sapien
2	3798	25.4	796	11 Q8CFX5	Q8CFX5 mus musculus
3	3445	23.0	803	11 Q8W0T2	Q8W0T2 mus musculus
4	2781.5	18.6	2649	5 Q9W092	Q9W092 drosophila
5	2775.5	18.5	2669	5 Q9W0T1	Q9W0T1 drosophila
6	2771.5	18.5	2669	5 Q95VB8	Q95VB8 drosophila
7	2705.5	18.1	645	11 Q8VDN7	Q8VDN7 mus musculus
8	1368.5	9.1	412	4 Q9H5E0	Q9H5E0 homo sapien
9	854.5	5.7	1711	5 Q45409	Q45409 caenorhabdi
10	848.5	5.7	1713	5 Q95Z08	Q95Z08 caenorhabdi
11	692.5	4.6	1022	5 Q960V3	Q960V3 drosophila
12	641	4.3	7210	5 Q9V7G8	Q9V7G8 drosophila
13	641	4.3	9270	5 Q9MLD9	Q9MLD9 drosophila
14	621.5	4.2	2768	5 Q9VC00	Q9VC00 drosophila
15	613.5	4.1	5476	5 Q9NJ17	Q9NJ17 drosophila
16	613.5	4.1	5533	5 Q9VP12	Q9VP12 drosophila

ALIGNMENTS

RESULT 1

Q9UG2	Q9UG2	PRELIMINARY;	PRT; 2781 AA.
ID	Q9UG2		
AC	Q9UG2		
DT	01-MAY-2000	(TrEMBLrel. 13, Created)	
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)	
DE	Bromodomain PHD finger transcription factor.		
DE	BTF.		
GN	Homo sapiens (Human).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606.		

RP SEQUENCE FROM N.A.
RX MEDLINE=20130111; PubMed=10662542;
RA Jones M-W, Hamana N, Shumate M, [REDACTED]
RT "Identification and Characterization BPTF, a novel bromodomain
RL transcription factor," [REDACTED]
RL Genomics 63(35-39)(2003): [REDACTED]
DR EMBL: AB032251; BAA89208.1; -.
DR HSP: Q92831; I891.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR004022; DDT_dom.
DR InterPro: IPR006209; EGF_like.
DR InterPro: IPR001965; Znf_PHD.
DR Pfam: PF00439; bromodomain; 1.
DR Pfam: PF02791; DDT; 1.
DR Pfam: PF00628; PHD; 2.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00571; DDT; 1.
DR SMART: SM00249; PHD; 2.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PSS00014; BROMODOMAIN_2; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PSS00016; ZF_PHD_2; 2.
DR PROSITE: 2781 AA; 311210 MW; 7B3BE0C1C4C3CFBE CRC64: 90

Query Match

94.6%: score 14170: DB 4: Length 2781:

Best Local Similarity 95.18; Pred. No. 0;		Matches 2766; Conservative 2; Mismatches 13; Indels 126; Gaps 1;	
Qy	1	MVSEEEEDGDAEETQDSEDEDEDEEDDDSDYPPEMEDDDDDASYCTESSFRSHST	60
Db	1	MVSEEEEDGDAEETQDSEDEDEDEEDDDSDYPPEMEDDDDDASYCTESSFRSHST	60
Qy	61	YSSTPGRKPRVHRPRSPILBEKDIPLPEPPKSSSEDLMPVNEHIMNVIAIYEVLRNFGTV	120
Db	61	YSSTPGRKPRVHRPRSPILBEKDIPLPEPPKSSSEDLMPVNEHIMNVIAIYEVLRNFGTV	120
Qy	121	LRLSPFREDECAALVSOEQCTLMAEMHVLLKAVLREEDTSNTFGPADLKDSVNSTLY	180
Db	121	LRLSPFREDECAALVSOEQCTLMAEMHVLLKAVLREEDTSNTFGPADLKDSVNSTLY	180
Qy	181	FDGMTPEVLRVYCESDKYHHVLPYQEAEDYPYGPVENKIKVLQFLVQOFLTNIARE	240
Db	181	FDGMTPEVLRVYCESDKYHHVLPYQEAEDYPYGPVENKIKVLQFLVQOFLTNIARE	240
Qy	241	ELMSGVIOYDDHRCVCHKLGLCCETCSAVYHLECVKPPLEVPDEMOCEVCVAHKV	300
Db	241	ELMSGVIOYDDHRCVCHKLGLCCETCSAVYHLECVKPPLEVPDEMOCEVCVAHKV	300
Qy	301	PGVTCVAEIOKNKPYIRHEPIGYDRSRKRYWFLNRRLIIEEDTENENEKKIWIYSTKVQ	360
Db	301	PGVTCVAEIOKNKPYIRHEPIGYDRSRKRYWFLNRRLIIEEDTENENEKKIWIYSTKVQ	360
Qy	361	LAELIDCLDKOYWEAELCKILBEMREETHRHMDITEDLTNKARGNKSFLLAANEIILES	420
Db	361	LAELIDCLDKOYWEAELCKILBEMREETHRHMDITEDLTNKARGNKSFLLAANEIILES	420
Qy	421	IRAKGIDIDNVKSPETEKDKKNETENDSKDAKNEEPEQOSLEKDSDDKTDDDPDEQK	480
Db	421	IRAKGIDIDNVKSPETEKDKKNETENDSKDAKNEEPEQOSLEKDSDDKTDDDPDEQK	480
Qy	481	SEPTVEVGDKGNSVANLGDNTTATSEETSPSEGRSPVGCLETSPDSSNMAEKKVASEL	540
Db	481	SE-----	482
Qy	541	PQDVPEPNKTCSSNTSATTTTSIQPNLENSSSSELNSSQSESAAKADDPENGERSHT	600
Db	483	-----	482
Qy	601	PVSIQEEIVGDFTEKSTGELSPGAGKGAGSGSTRIITRLRNPDSKLSQLKSQOAAAA	660
Db	483	-----VGDfKSEKSNSELSGSPGAGKGAGSGSTRIITRLRNPDSKLSQLKSQOAAAA	534
Qy	661	HEANKLFKEGKEVLVYVNSOGEISRLSTKKEVIMKGNINNYFKLGQEGKYRVYHNOYSTNS	720
Db	535	HEANKLFKEGKEVLVYVNSOGEISRLSTKKEVIMKGNINNYFKLGQEGKYRVYHNOYSTNS	594
Qy	721	FALNKHQHRDHDKRHLAHKFLCTPAGFEKNGSVHSGSKVLTISTLRLTIQLENNIPS	780
Db	595	FALNKHQHRDHDKRHLAHKFLCTPAGFEKNGSVHSGSKVLTISTLRLTIQLENNIPS	654
Qy	781	SFFHPNASHRANWIKAVQMCSPREFALALALECAVRPVVMLPTWRFLGHTRLHRMT	840
Db	655	SFLHPNASHRANWIKAVQMCSPREFALALALECAVRPVVMLPTWRFLGHTRLHRMT	714
Qy	841	SIBREEKVKKEKKEEETMQOATWVKYTFPVKHQVWKQKGEYRVTYGGWSWISK	900
Db	715	SIBREEKVKKEKKEEETMQOATWVKYTFPVKHQVWKQKGEYRVTYGGWSWISK	774
Qy	901	THVYRFYKPLPGNTNVNYRKSLEGTKNMNMENDESDDRKRCSRPKKIKIEPDSKDEVK	960
Db	775	THVYRFYKPLPGNTNVNYRKSLEGTKNMNMENDESDDRKRCSRPKKIKIEPDSKDEVK	834
Qy	961	GSDAAKGADQNMEDISKITEKQDQVKELLDSDSKPCKEEPMEDDDMKTESHVNCQES	1020
Db	835	GSDAAKGADQNMEDISKITEKQDQVKELLDSDSDKPCKEEPMEDDDMKTESHVNCQES	894
Qy	1021	SOVDVNVSEGFHLRTSYKKTKSSKLDGLLERRIKOFTLEEKQRLKIKLEGGIKGK	1080

Db	895	SOVDVNVSEGFHLRTSYKKTKSSKLDGLLERRIKOFTLEEKQRLKIKLEGGIKGK	954
Qy	1081	TSNSSKNLSSESVITKAKEGCQSDSMRQEQSPNANNQPEDLIOCCSOSDSVLRMSDP	1140
Db	955	TSNSSKNLSSESVITKAKEGCQSDSMRQEQSPNANNQPEDLIOCCSOSDSVLRMSDP	1014
Qy	1141	SHTTNKLYPKDRVLDVDSIRSPETKCPKONSTENDIEEKVSOLASRGQPTKSKTKGNDP	1200
Db	1015	SHTTNKLYPKDRVLDVDSIRSPETKCPKONSTENDIEEKVSOLASRGQPTKSKTKGNDP	1074
Qy	1201	FIDDSKLASADDIGTLICKNKPLIOEESDTIVSSSKSALHSSVPKSTNDROATPLSRAM	1260
Db	1075	FIDDSKLASADDIGTLICKNKPLIOEESDTIVSSSKSALHSSVPKSTNDROATPLSRAM	1134
Qy	1261	DFEGKLGCSSESNTLENSSDTVSIQDSSEEDMIVQNSNESISEQFRTREQDVEVLEPLK	1320
Db	1135	DFEGKLGCSSESNTLENSSDTVSIQDSSEEDMIVQNSNESISEQFRTREQDVEVLEPLK	1194
Qy	1321	CELVSGETGNCEDRLPVKGTANGKKPSQOKLEERPYNKCSDOIKLKNTTDDKKNNENR	1380
Db	1195	CELVSGETGNCEDRLPVKGTANGKKPSQOKLEERPYNKCSDOIKLKNTTDDKKNNENR	1254
Qy	1381	ESEKKGORTSTFOINGKDNKPKIYLKGECLKEISESVVSGNVEPKVNNINKIIPENDIK	1440
Db	1255	ESEKKGORTSTFOINGKDNKPKIYLKGECLKEISESVVSGNVEPKVNNINKIIPENDIK	1314
Qy	1441	SLTVKESAIRPFIINGDVIMEDFNERNSETSKSHLLSSDAEGNRSLETLPTSKESDST	1500
Db	1315	SLTVKESAIRPFIINGDVIMEDFNERNSETSKSHLLSSDAEGNRSLETLPTSKESDST	1374
Qy	1501	QTTTPSASCESNSVNOVEDMETSEVKKVTSPTSEESNLNSDFIDENGLPINKNE	1560
Db	1375	QTTTPSASCESNSVNOVEDMETSEVKKVTSPTSEESNLNSDFIDENGLPINKNE	1434
Qy	1561	NVNGESKRKTIVTEVTMTSTVATESKTVIKVEKGQKQTVVSVSTENCAKSTVTTTTVT	1620
Db	1435	NVNGESKRKTIVTEVTMTSTVATESKTVIKVEKGQKQTVVSVSTENCAKSTVTTTTVT	1494
Qy	1621	KLSTPSTGGSVDDIISVKEQSKTVTTTDSLTGTLTMTVSKYESTRDKVKLMKF	1680
Db	1495	KLSTPSTGGSVDDIISVKEQSKTVTTTDSLTGTLTMTVSKYESTRDKVKLMKF	1554
Qy	1681	SRPKKTRSGTALPSYRKFPVTKSTKKSIFVLPNDLKKLARKGGIREVPYFNYNAKPAIDI	1740
Db	1555	SRPKKTRSGTALPSYRKFPVTKSTKKSIFVLPNDLKKLARKGGIREVPYFNYNAKPAIDI	1614
Qy	1741	WPYPSRPTFGITWRYRLQTVKSLAGVSLMLRLWLASLRWDDMAAKVPPGGSGSTRTETSE	1800
Db	1615	WPYPSRPTFGITWRYRLQTVKSLAGVSLMLRLWLASLRWDDMAAKVPPGGSGSTRTETSE	1674
Qy	1801	TEITTTTEILIKRQDVGPYGIREFEYCIKIKIPIGVPETPKETPTQPKGLRSSALRPKRPE	1860
Db	1675	TEITTTTEILIKRQDVGPYGIREFEYCIKIKIPIGVPETPKETPTQPKGLRSSALRPKRPE	1734
Qy	1861	TPKQTPGVIIETWAAEELEWEIRAFARVEKEKAQAVEQQAQKRLQEQKPTVIATSTT	1920
Db	1735	TPKQTPGVIIETWAAEELEWEIRAFARVEKEKAQAVEQQAQKRLQEQKPTVIATSTT	1794
Qy	1921	SPTSSTTSTISPAQKVMVAPISGSVTTCTKMVLTTKVGSPATVTFQONKNEHQTFATWVK	1980
Db	1795	SPTSSTTSTISPAQKVMVAPISGSVTTCTKMVLTTKVGSPATVTFQONKNEHQTFATWVK	1854
Qy	1981	QGSQNSGVVQOQKVLGIIPSSTGTSTSOQFTTSFQPRTAIVTIRPNTSGSGGTTNSQVIT	2040
Db	1855	QGSQNSGVVQOQKVLGIIPSSTGTSTSOQFTTSFQPRTAIVTIRPNTSGSGGTTNSQVIT	1914
Qy	2041	GPQIRPGMTVIRTPLOQSTLGAIIIRTPVMVQPGAPQVMTQIIRGQPVSTAVSAPNTVS	2100
Db	1915	GPQIRPGMTVIRTPLOQSTLGAIIIRTPVMVQPGAPQVMTQIIRGQPVSTAVSAPNTVS	1974
Qy	2101	STPGOKSLTSATSTNIOSSASOPPRPOGOQVKLTMAQITQLTQGHGNGOGLTVVIOGOG	2160
Db	1975	STPGOKSLTSATSTNIOSSASOPPRPOGOQVKLTMAQITQLTQGHGNGOGLTVVIOGOG	2034


```
QY 1483 -----NYRDSL-----ETLPSK-----ESDSTQTTTPSASPESYNNOVEDM- 1521
Db 1060 ARAHAVPFDLSLLFECILVQOECVNTIKIQEYNASSGCNTTDSNTQDSKIDYIESMD 1119
QY 1522 -----EETSVMKVTSPITSESNL-----SNDFID-----ENGLP 1555
Db 1120 VCSNVEIESTEDSVTGLNSGNAEDVMTPCWRRKRKNQSKSYIGTKDVLDTQDLKDIP 1179
QY 1556 INKNENVNESKRKTVTEVTMTSTVATESKTVIKVEKDGKQTVWSSTENCASVTYTT 1615
Db 1180 LNK-----QNRFFPII-----ARPKREC-----VKYERETFNGERNV----- 1214
QY 1616 TTTVTKLSTPGSGSV-----DIISVKEQSKTVTTTVDLSLTGTGGTLVTSMTVKEYST 1671
Db 1215 -----YSTSPRGVYLLNDAKLYEQ-----VK 1240
QY 1672 RDKVKLMFRPKKTRSGTALPSYKFKVTSTKTSIFVLNDDLLKLARKGGIREVPYFN 1731
Db 1241 EDKSTIK- -KPSYSR-----YPLISNFTLHKKRSLLVLPREFLLKLARLGGKSTNGFH 1294
QY 1732 YNAKPAIDWPPSPRPTFGITWRYRLQTVKSLAGVSLMLRLLWASLRWDDMAKVPBG 1791
Db 1295 HAAKNN-TIWQYQSRPLFTCWSYRTSNATSLSLALQKLWLSCLRWDMDIAK-PPST 1352
QY 1792 GSTRTETSETITTEIIRKRDVGPYIRFEYCIKRIICPIGVPTPKETPTPQKGLRS 1851
Db 1353 DGKHQVTTDTTEIVTLELLKLRHSGRYGKTSYLRKVVIPLEMPKTVREV-TSIRSLR- 1410
QY 1852 SALRPKRPETPKQTPGVITETWABEELEWEIRAFERKEKAQAVEQO-AKKRLEQ 1910
Db 1411 ---KRAESPOPTPEITEWVDEDKLEIKRPMGEKQKARLSAVTRSVASRQLE-- 1465
QY 1911 KPTVTATSTSTSTSTISPAQKVMYAP-IGSVTTTGTKMVLTKVGSPTATVFOCNK 1969
Db 1466 -----ASGNSGNTSTNGALGVAGVQLAPKLSEDVK---EKMEQQLK---QRAVHOQR 1515
QY 1970 NEHQTFATVWVGQSGNSGVVQOQKVLG-----IIPSTGTS-----QQTFTSFQ-----PRTA 2018
Db 1516 -----LVATGEITRSVTPVKGVIGSRRVIVKPNPDGTTRIQQAVTVQVSRGTGANTA 1567
QY 2019 TVTIRNTSGSGGTSNS-----QVITGPQ-----IPGMTVTRTP-----LOO 2057
Db 1568 AAAASPTVGGSTSTQSNPSTSTPHKVQIIRGDPGVSRGLNPGQOLVOMPDGKLHLVTT 1627
QY 2058 STLGLKALIRTPVMVOPGAPOQVMTQIIRGOPVSTAVSAPNTVSTPGOKSLTSTATSTNI 2117
Db 1628 TTSSNSAGOGNKKVPKIPASTSS-----SPAISSAQTTNPVTPVVIKQIAVHVTKNSAT 1683
QY 2118 QSSASOPPRPQOQGVKLTMAQLTQTHGGNGQLTVVVIQOGQTTGQ-----LQLI 2169
Db 1684 QSTAS-----SSRVALPLAQI-----KNKLLAQOQOQSTSSSPATSSSPVQKI 1727
QY 2170 PQGVTVLPGGQQLMAAMPNGTVQRFLETPPLATTATTASTTTTSTVTA--GTGEOR 2226
Db 1728 VSKVWNTSTSGTLOQVFOGSS-----KLWVQNAQOGKVIIISTSAAOQOQTSVPQ 1779
QY 2227 QSKLSQMOVHODKTLPPAQSSS-----VGPAK--AQPTAQPSARPOQTPQSPAPQ 2278
Db 1780 QOQLVQSQPIQS-----PQISMTQOQIVVGGRIILSPGQITVITQNVNVPQSQALQWQO 1835
QY 2279 EVGTQP-----EVQOTQTVSSHVPSAQPTHQAQSSKPPQVAAGSQPSNVQOGSPVRV 2330
Db 1836 QIQTOQOQOQHVVQPQOQFVVGQSNQIVGSSPSAOTKLKQLVWQOQSOOTIEETQIIT 1895
QY 2331 QSPSQT-----RTRPSTPQLSPGQOSQOVQTTTSQPIPIOPHTSLQ-----IPSGQ-- 2377
Db 1896 TDSNETGTQOVLVNPSTLTAQQLAQGLQVATVNGQOQVIVKPLGNNAQVIAHIKHQGDGN 1955
QY 2378 -----POSQPOVQSSQTLSGGQTLNOVSVSSPSRPLQIQOQPOP----- 2417
Db 1956 AHIVTSNATVQANPQTSVPVKQOALPQSPQOVVQ-----QQQIHQGSPTNFBEGVT 2010
```

```
QY 2418 -----QVIAPVLOQOQOVV-LSQIOSQ-----VVAQIQAOQSG----- 2449
Db 2011 PITQOPVLTAQVAPQAQOQALSVEESLLQNPQPGTVIKCVTAQVLTQTEHGPRIVLOGLVG 2070
QY 2450 -----VPQOIKLOLP-----LOIQOSSAVOT----- 2470
Db 2071 NDFTAQOQLQVOTQVQOQLMKQAQESNGKLVGTPKTIYLAOPENAVQSPPLTPVHQS 2130
QY 2471 --HOIQNVTVQA---ASVQEQQLRVQQLR-----DQOQKKQOQIEINVTNSKL----- 2516
Db 2131 AAHQQTNNIETADTLATTYEANSTIKDIAINNGDDOENSKCAETENSITNESFAGTS 2190
QY 2517 -----LIKVEIQOVVMKHA--VIEHLKOKKSWTPAEREEN-----QRMI 2556
Db 2191 SILLEGSEHDEPTNLAGLIDISETDLENKQNESFVVTTRGYTOKSISNALKQNLISPELEKL 2250
QY 2557 VCNQVMKYIILDKTDEE-----KQAAKKRKEESVEOK----- 2589
Db 2251 VCMQKQOENANSTNETCSRGVNEALTPSRQTDDETKITSLRRPNAMTTSSQFNR 2310
QY 2590 -----RSQN-----ATKLSALLFKHKEQLRABELKKRALLDKDQIQIEVQOELKDLKTK 2639
Db 2311 ILKKNRSKNDEVAELGEGQOSQLERHKKELLLKKNILKRSLLERNLQSEIHE-----DVKT 2366
QY 2640 KEKDLMLAQATAVAPCPVTPVLPAAPPSPPPPPGCVQHTGLLSTPLPV----- 2693
Db 2367 VORHVRPLSNA-----SPDEQSENERSG---EPNLDKFKTEVQ 2401
QY 2694 -----ASOKRKRREEK-----DSSSKSKK 2713
Db 2402 NPHGAGRPKLLTRKKEKLYCICRTPYDDTKFYVGCDCLSNWFHGDVCSITEASK-KLS 2460
QY 2714 KMISTTSKETKDKTKLYCICKTPYDESKFYIGDRCONWTHGRCVGIQASEALIDYVC 2773
Db 2461 EFICIDCKRARETQQLYCSQRPYDESOFYICDCKODWPHGRCVGIQASEAFIDEYVC 2520
QY 2774 PQOSTEDAMTV-LTPLTEKDYELKRVLSLOAHKMAWPFLEPVDPNADPYGVVKEP 2832
Db 2521 PECORNDANANANKKLTSTNDVEELKNIKOMQLHKSANWFMEPVDPEADPYKVKEP 2580
QY 2833 MDLATMEERVQRRYYEKLTEFVADMTKIFDNCRYNPSDSPPFYOCAEVLSFVQVKLKG 2892
Db 2581 MDLKRMEIKLESNTYTKLSEFIDGTMKIFDNCRYNPKESFYKCALESEYFQKIKNF 2640
QY 2893 K 2893
Db 2641 R 2641

RESULT 5
Q9W0T1 PRELIMINARY; PRT: 2669 AA.
AC Q9W0T1;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE CG32478 protein.
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
```


Db 631 AASFLNINVVNKKIWAANVARRPSEFAVALLLFOASLKSUVFANVWHEQLGHTTLQR 690
QY 839 MTSIREBEKEVKKKEKQ---EBEETMQOATWVYFPVKHVKWOKGGEYRVYTGCGW 895
Db 691 ITSAREBERKKLEKREKREDEBERNLAENYIKYTLGLKHVKWOKGGEYRVHGWGOW 750
QY 896 SWLSKT---HVRVFPKLPNGTWNVYKRSLEGTNNMDENMDESCKRCSRPPKIKI 950
Db 751 LMLSSRRCGVARRAQPLTNRVYVHTM---GEENDVNEII-----LVDPRTORF 799
QY 951 EPDSKDEVKSDAAKAGDQ--NEMDISKIITEKKDQDVKELLDSDSKPCKEEMPEVDDDD 1008
Db 800 MQOCESSNVGQVCHYLPDQYKNVKVEDVTE----- 831
QY 1009 MKTESHVANCQESSQVDVNVNVEGHL--RTSYKKKTKSKLDGLLERRIKOFTLEEKORL 1066
Db 832 -KIGH-----IDVSKALNAPGRTYISKVARKSRJDDDLDRKLKLAEEV--QMA 878
QY 1067 ERIKLEGGIKIGIKTSTNSKNLSESPVITKAKEGQSDSMRQEQSPNANNQDPEDLIQ 1126
Db 879 SKI----- 881
QY 1127 CSQSDSSVLMSDPSSHNTNKLKPKDRVLDDVSIIRSPETKCPKQNSIENDIEEKVSDLASR 1186
Db 882 --PSDMKPLLVSSQNTANS-----KOTFLE-----KR 907
QY 1187 GOEPTKSKTKGNDFIDDSKLASADDIGTLICKNNKPLIQESDITVSSKKSALHSSVPK 1246
Db 908 LRLTEVQAKG----- 918
QY 1247 STNDRDATPLSRAMDFEGLKGCDSNSTLENSDTVSIQDSSEEDMIVQNSIESIQF 1306
Db 919 -----GPANVNLEVLNSAKIQOTVRLOFS----- 943
QY 1307 RREQDVEVLEPLKELVSGESTGNCEDRLPKVGTTEANGKPSQOKLEERPKNKCSDOI 1366
Db 944 ---QLNRFKAFVRC---YKEC-----NTNSNAVSQITO-----NFCYSLP 978
QY 1367 KLKNTTDDKN---NENRESEKKGQRTSTFQINGKDNKPKIYKLGECLEKISESVWSGN 1422
Db 979 CLQKARAKELLLLRKAHTAGNAGSKETVAAIILGAVKKPSIL-----EQKLTEGK 1028
QY 1423 VEPKVNINKIIPENDIKSLTVKESAIRPFINGVIMEDFNERNSETKSHLLSSSDABG 1482
Db 1029 RE-----STQAVAVDDSEEGKPAESEAPLDLLLDQDWEH 1059
QY 1483 ---NYRDSL-----ETLPSTK-----ESDSTOTTTTPSASCPSNSVNOVEDM- 1521
Db 1060 ARAHAVPFDSSLTECILLVDQECVNTKIKQEVNASSGCGNTTPDSNTQDSKDIDYESMD 1119
QY 1522 -----BIETSEVKKVTSSPITSEESNL-----SNDFID---ENGLP 1555
Db 1120 VCSNVEIESTEDSIVTGLNSGNAEDVDMTPGWRKRNRNOKSKSYIGTKDVLDTOLDKIDP 1179
QY 1556 INKNENVGESKRKTIVTEVTMTSTVATKESKVIKVEGDKQTVSVSSTENCAKSTVIT 1615
Db 1180 LNK-----QNRFPFIT-----ARPVKREC-----VKKYRETFENGNERV- 1214
QY 1616 TTTVTKLSTPSTGGSV---DIISVKEQSKTVVTTVTDTSLATTGTTGTLTSMVSKEXST 1671
Db 1215 -----YSTSSPGRVILLNDAAKLYEA-----VKT 1240
QY 1672 RDKVKLMKTSRPKKTSGFALPSYKFKVTYKTSKISFVLPNDLKKLARKGGIREVPYFN 1731
Db 1241 EDRKSTITK--KPSYSR-----YPLISNFLTTHKKRSLVLPRLPELLKLARLGGKSSNTGNFH 1294
QY 1732 YNAKPALDIWYPSPPTGCTWRYELOVQVSLAGVSLMLRLIWLASLRDWDMAAKVPPGG 1791
Db 1295 HAAKNNTWQYQCSRPLFTCMYSYRTSNATLSSSLAQRLQLWNSCLRWDMDIAK--PPST 1352
QY 1792 GSTRTETSEITTEIIRRRDVGPGYCFIRFECIRKCIICPGIVGPPETPTPPQRGLRS 1851
Db 1353 DGKHQVYTDTEIVTEILLELLRHSGRYGKTSYLRKKRVWIPLEMPKTVREV-TSIRSGLR- 1410

QY 1852 SALRPKRPEPKOTGPVIIETWVAEBELELWEIRAFARVERKEKAQAVEQ--AKKRLQEO 1910
Db 1411 ---KRRAESPQTEPQITTEWVDEKLELWEIKFMGEQEKARLSAVTRSASRQLE-- 1465
QY 1911 KPVIATSTSPSSSTSTISPAQKVMVAP-IGSVTTTGTWKVLTITTKVQSPATVITQOKN 1969
Db 1466 ---ASGNSGNTSTNGALVAGRVOLAPKLSEDKV--EKMBEQQLKL--ORAVHQOR 1515
QY 1970 NFHOTFATWVKOGSNGVVQOKVLG---IIPSTGTS---IIPSTGTS---OQTFTSFQ---PRTA 2018
Db 1516 ---LVATGEITRSVTPVKQVIGSRVIVKNPDGTTRIIOQAANTOVSTGANTGA 1567
QY 2019 TVTIRPNTSGSGGTSNS-----QVIRGPQ-----IRGPMVIRFP-----LQO 2057
Db 1568 AAAASPVTGSGTSTOSNPSTSTPHKVQIIRGPGKVSVRGLNPGQQLVQMPDGLHLVLT 1627
QY 2058 STIGKALIRPVAVQCAPQOVMTIIRGQPVSTAVSAPNTVSTTCKOKSLTSTSTNI 2117
Db 1628 TTSNSAGOKMKVPKIPASTSS-----SPAISSAQTTNPVTPVVIKQIAVKHVTNKSAT 1683
QY 2118 QSSASQPRPQOQGVKLTMAQLTQTOGHGNGOGLTVYIOGOGQTTGQ-----LOLI 2169
Db 1684 QSIAS-----SSRVALPLAQI-----KNKLLLAQOQOQSTSSSPATSSSPVKI 1727
QY 2170 PQGVTVLPQCGQQLMQAMPNGTVORFLPTLATTATTATTTTIVTSTAA---GTGEOR 2226
Db 1728 VSKVYNTSTSGOTLQOVFVQSGS-----KLVGQNAQOGKVIIISTSAAQOQGTSPVQ 1779
QY 2227 QSKLSPQMVHODKTLPPAOS--SVGPAKAQ-----PQ 2258
Db 1780 QOQLVQSQPQOOS-----PQOISMVQVGNQPTQKVIOIVNTSNVQOQIIVGGORILLSPG 1835
QY 2259 TAQPSARPQOTQPSQAQPEVOTQP-----EVOTQTTVSHVSEQAOPHAQSSKP 2310
Db 1836 QTIIVTORNVQSOALQMVQOQIQOQOQHHVQPOQOQFVVOQNOIVQSSPSAQTKLVK 1895
QY 2311 QVAASQPOSQVQSPRVQSPSQT-----RIPSTPSQLSPQSQSOVOTTTTSQPIQ 2365
Db 1896 QLVVQOQSQQTIEKTOITTTDSNETGTQOVLPNPNSTLAOLAQGLQVATVNGCQVIVK 1955
QY 2366 PHTSLQ-----IPSQGO-----POSQPVQSSQSTQTLSSGOTLQNVSVSSP 2405
Db 1956 PLGNNAQIVAHIKHQGDGNAHIYVNSATVPAQNPQTSVPKQOALPPQSPQOVVVO-- 2013
QY 2406 SRPOLQIOQPOP-----QVIAPVQLOQOVV--LSQIOSQ-----V 2439
Db 2014 ---QOOIHQOQSPTNFESGVTPITQOQVLTQAVQAPAOQOALSVEESLLQNPQPTVIKCV 2070
QY 2440 VAOIOAQOQSG-----VPOQIKLQLP-----IQ 2461
Db 2071 TAQVLQTEHGPRIVLOGLVGNDFTTAQOQLVOTQVKOOLMKAQESNGKLVGLGPKIYLA 2130
QY 2462 IQOQSAVOT-----HQIQNVVTVQA---ASVQEQLOQVQOQLR-----DQOQK 2500
Db 2131 VOPENAVQSQPPLTPPVHOSAAHQOTNNIEADTTLATTYEANTSTIKDIANNNGDQENS 2190
QY 2501 KQOOQIEINVTPSKL-----LIKVEIIQOVVMKHNA--VIEHLKQK 2541
Db 2191 KCAETPENSNTITNESFACTSLLGSEHDEPTNLGLDIDSETDLENKQNESFVTVIRGVIQ 2250
QY 2542 KSMTPAEREEN-----QRMVNCQVMKYILDKIDKEE-----KQAAKRRKREESVEQ 2588
Db 2251 KSIISNALQOGLNLSPELEKLVCMQKQOENANSTNEWETCSRGSVNEEALTPSRQTDDEW 2310
QY 2589 K-----RSKQK-----ATKLSALLFKHQLRAELIKKRAL 2619
Db 2311 KIRTSILRRPNAMTTSSQFNRIILKKNRSKNDEVAELGEQKQSLERHKKLLKNILRRKSL 2370
QY 2620 LDKDLOIEVQEBELKPDALKKKEKDLQMLAQATAVAAPCPPVTPVLPAPPAPPSPPPPG 2679
Db 2371 LERNJOSEIHE-----DVTKVQVRHVRPLNSA-----SPDEQSE 2404


```

Db      882 --PSDMKPLLYSSQNTANS-----KQTFLE-----KR 907
QY      1187 GOEPTKSTKNDFFIDDSKLASADDIGTLICKNKKPLIQEESDTIVSSSKSALHSSVPK 1246
Db      908 LLRLTEVOAQG-----
QY      1247 STNDRDATPLSRAMDFEGLGCDSESNTLSSSDTVSIQDSSSEEDMIVQNSNESISEQF 1306
Db      919 -----GPANVNLLEVNSLAKQIQTVRLQFS-----943
QY      1307 RTREQDVEVLPLKCELVSGESTGNCEDRLPVKTEANGKPKSOOKLEERPVNKCSDQI 1366
Db      944 ---QLNFAKVCRC-----YTKEC-----NTNSNAVSIQITQ-----NTCYSPL 978
QY      1367 KUKNTTDKN-----NENRESEKKQORTSTFOINGKDNKPRIYLKGECLKEISESRVSGN 1422
Db      979 CLQKARAKKELLLLRKAHTAGNSKETVAAILGAVKKPSIL-----EOKLTEGK 1028
QY      1423 VEPKVNINKIIPENDIKSLTVKESAIRPFINGDVIMEDFERNERNSETKSHLLSSDAEG 1482
Db      1029 RE-----STQVAVDDSEEGKPAEAEPLDLLLDQWEH 1059
QY      1483 -----NYRDSL-----ETLPSTK-----ESDSTQTTTPSASCPSNSVNOVEDM- 1521
Db      1060 ARAHAVPFSDSLLTECILLVQCEVTNTKIQEVNASSCNCITPDSNTQDSKDIDYIESMD 1119
QY      1522 -----EIEISEVKKYTSPTISEESNL-----SNDFTD-----ENGLP 1555
Db      1120 VCSNVEIESTESIVTGLNSGNAEDVMTPGWRRKRNOKSKSYIGTKDVLQDLDKQIP 1179
QY      1556 INKNENVNGESKRKVTIETVTMTSTVATESKTVIKVEKDKQTVYSSSTENCAKSTVTTT 1615
Db      1180 LNK-----QNRFPIT-----ARPVKREC-----VKKYERETPENGERNV----- 1214
QY      1616 TTVTVKLTSPSGGV-----DIIISVKEOSKIVVTTVTDLSLTGTGLVTSMTVSKEYST 1671
Db      1215 -----YSTSPSRGRVYLLNDAAKLYEA-----VKT 1240
QY      1672 RDKVKLMKSRPKTRSGTALPSYRKFTKSTKSFIVLPNDLKKLARKGGIREVPYFN 1731
Db      1241 EKSITTK--KFSYR-----YPLISNFLTHKKRSILLVLPRELLKLARLGKSKSTNGFH 1294
QY      1732 YNAKPALEDIWPYSPRPTEGIIWRYRLQTVKSLAGVSLMLRLWLASLRWDDMAAKVPVGG 1791
Db      1295 HAAKNN-TIWQYQCSPLPRTCSYRTSNATSLSSALQLRLWLSCLRWDMDIAK-PPST 1352
QY      1792 GSTRTTSETTEITELIKRRDVGPIRFEYCIRKICPIGVPTPKETPPORKGLRS 1851
Db      1353 DGKHQVTTDTEIITLELLKLRHSRGYKTSYLRKRVIPLEMPKTVREV-TSIRSGLR- 1410
QY      1852 SALRPRRPETPKOTGPVIIETWVAEELELWIRAFARVEKEKAQAVEQO-AKKRLEQO 1910
Db      1411 ---KKRAESPQTEQITEEWNVEDKLELWEIFKMGKEKARLSAVTRSVASRQLE-- 1465
QY      1911 KPTVIATSTSPSTSTTSISPAQKVMWAP-LSGSVTTCTKAVLTIKVGSPTATVFOQNK 1969
Db      1466 ----ASGNGSNTSNGALGVAGRVQLAPKLSYEDV--EKMEQOLKL---QRAVHOQRK 1515
QY      1970 NFHTQATWVKQSGNSGVVQOQVLG----IIPSTGTS---QOTFTSFQ-----PRTA 2018
Db      1516 -----LVATGEITRSTVPKGVIGVRSRIVRNPDPGTTRIIOQAVTQVSRGTGANTA 1567
QY      2019 TVTIRNTPSGSGTTSNS-----QVITGPQ-----IRPGMTVIRTP-----LOO 2057
Db      1568 AAAASPVTGSGTSTQSNPSTSTPHKVOIIRGPDGKVSVRGLNPGQOOLVQMPDGLHLVJT 1627
QY      2058 STLGKALIRTPVMVQCAPQOQWTOIIRGQPVSTAVSNAPTYSSTPGOKSLASATSNM 2117
Db      1628 TTSNSAGQGNKMKVPIKPASTSS-----SPAISSAQTTTNPVTPIQIAVXHVTKNSAT 1683
QY      2118 QSSASQPPRPQOQKLTWALQTLTQGHGNGGLVAVIQQGQQTGTQ-----LQLI 2169
Db      1684 QSIAS-----SSRVALPLAQI-----KNKLLAQOQOQSTSSSPATSSSPVQKI 1727

```

RESULT 7

Q8VDN7

ID Q8VDN7

PRELIMINARY;

PRT; 645 AA.

Db	363	PAASQKREKSSSKKKMLSTSTSKETKDKTKLYCKTYPYDESK	412
RESULT 9			
ID	045409	PRELIMINARY; PRT; 1711 AA.	
AC	045409;		
DT	01-JUN-1998 (Tremblrel. 06, Created)		
DT	01-JUN-1998 (Tremblrel. 06, Last sequence update)		
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
DE	F26H11.2a protein.		
GN	F26H11.2 OR F26H11.2A.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Barlow K.		
RL	Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99069613; PubMed=9851916;		
RA	none;		
RT	"Genome sequence of the nematode C. elegans: A platform for		
RT	investigating biology."		
RL	Science 282:2012-2018(1998).		
DR	EMBL; Z81515; CAB04197.1; -		
DR	WormPep; F26H11.2a; CE15908.		
DR	InterPro; IPR000637; AT.hook.		
DR	InterPro; IPR004022; DDT.dom.		
DR	InterPro; IPR001965; znf_PHD.		
DR	Pfam; PF02791; DDT; 1.		
DR	Pfam; PF00628; PHD; 1.		
DR	SMART; SM00384; AT.hook; 1.		
DR	SMART; SM00571; DDT; 1.		
DR	SMART; SM00249; PHD; 1.		
SQ	SEQUENCE 1711 AA; 196549 MW; AC45F73FD894044E CRC64;		
Query Match 5.7%; Score 854.5; DB 5; Length 1711;			
Best local similarity 17.3%; Pred. No. 3.2e-28;			
Matches 431; Conservative 296; Mismatches 624; Indels 1139; Gaps 72;			
Qy	4	EEEEEE---DGAETQSDDE---EDMEEDDDSDYPEEMEDD---DDDAASYCTE	52
Db	169	KROEEDIIYDEDESEEESSDDEFMLNEDQVVQEEELNLTDIKIEGLDEENKYC--	226
Qy	53	SSFRSHSYSTPGRRKPRVHRSPILIEE--KDIPPLEFPKSSDLMVPNEHIMVIAI	110
Db	227	-----PWLDEDPAFLPKLEPSSQDIPPTASIMDAVEI	261
Qy	111	YEVLNFGIVRLSPREFDFCAALVSOQCFTLMAEMHVVLKAVLREEDTNTTFFGPAD	170
Db	262	YEILRSYHRTLRITPTTFEDFCAALISHNNSCIMAEVHALLRNCLKSDDEQTHSYTE	321
Qy	171	LKDSVNSTLYFIDGMTWPEVLRVYCESDKYHHVLPYQEA-----BDYPGPVENKI	222
Db	322	TNNSVNIIMHMDTLVYAEILRQYIEA-----YFADASVRDAINVDNYPFVGVDANI	374
Qy	223	K-----VLOFLVDQFLTNIAREELMSEGVIOYDDHCRVCHK-LGDLLC	265
Db	375	QRDFSEFFNKHVFRVLWFLMSYRELYSFEKFLVNNYVGFONDENCRVCGKSRVVG	434
Qy	266	CETCSAVYHLEGVKPPLEEVPEDEMOCEVCAVHK-VPGVTDVCVAEQKNKPIRHPICY	324
Db	435	CTOCEAAHFVHC--SHLKFPPE-VIVCNICKNSAVRGVLPDEADVREP--LRSQPIGR	489
Qy	325	DRSRKYWFNLRIIEEDTENENKKIWIYTKVOLAEILDCLDKDYWEAELCKTLEEM	384
Db	490	DRYGRYWFVIRLVVQSLDETE---LYYSTVPQLYQLQKLDRTYYEKDLCDTIRLR	545
Qy	385	RELIHRMDITDLTNKAGSNKSFLLAAANEELIESIRAKK---GDINDVKSPEETEKDK	441
Db	546	IDFLEQMALTVEMTSERR-----EAALETMVKRQLIGYDFAEATTPQ-----	588
Qy	442	NETENDSKDAEKNEFEEDOSLEKSDDKTPD--DDPEOGKSEEPTEVGDCKGNSVSNL	498
Db	589	-----IYLHRDSMKRMASILRDCAQKGQVKQEVKL-----	618
Qy	499	GDNTNATSEETSPSEGRSPVGCISETPDSSNMAEKKVASELPQDVPEEPNKTCESSNTS	538
Db	619	-----EE--PVEGSPVKVCQFVEDSI-----LPE-----	641
Qy	559	ATTTSIQPNLNSNSSELSSQSESAKAADDPENGERSHTPVSIOEIVGDFTSKST	618
Db	642	-----SMIGIFDAKLINTFWSGGA-----TOEIVVEQFV---DI	672
Qy	619	GELSESPCAGKAGSGSTRITRLRNPDSKLSQKSOQVAAAHAANKLFKEGKREVLVNS	678
Db	673	SDNFDAPSA-----	681
Qy	679	QGEISRLSTKKEVINMGNNINNYFKLQEGK---YRVYHNOYSTNSFALNHQHREDHDKR	735
Db	682	-----NLWRMGDEGNDQTFMTYTYNYSRNMESSEFLTRKKAADKK	721
Qy	736	RHLAHKFCCLIPAGEFKW---NGSVHGSKVLTISTLRLTITOLENNIPSSFFHPNASHR	791
Db	722	KYMAKSF--AQIDNEDVWVAKNQFYGDASHCKFIMWTLOQVYIKNIPIDLHMKKWEFA	779
Qy	792	ANWIKAVQMCSPREFALALALECAVKKPVVMLPIWREFLGHTLRLHMTSIEEREKVKV	851
Db	780	KGFDLEVSADDDYKLVCTLLKLDCAVKKTFIMFQWNGLGQTRLEITVDQRENFKEQ	839
Qy	852	KKEKK-----QREETMOQATWVKYTFPVKHQVWKQKGEERYVTG---YGGHWSLKT	901
Db	840	QRLKKIDADALTLDLDDSFVRVNYMKPKWPTY-ILRQGETYRNAGKSGMGWAWAAK	898
Qy	902	HVYRFVPLPGNTVNVYRKSLGEGTKNMNDENMDSKRSRSPKKIKIPDSEKDEVKG	961
Db	899	YVEKWI-----QVPSPKPLAVTVEELKTESVNRK---930	
Qy	962	SDAAKGADQNEWDISKITEKKDQDVKELLDDSDSKPCKEPEMVEDDMDKTESHVNCQESS	1021
Db	931	-----ARRLELVSKIT-----	942
Qy	1022	QDVVNVSEGPHLTSTYKTKYKSKLDGLLERRIKQFTLEBKQRLKIKLGGIGKIGTK	1081
Db	943	-----KKQRSG-----GKS	952
Qy	1082	STNSSKNLSESPVITKAKEGCQSDSMROEQSPNANNQPEDLIQGCQSDSSVLMSDPS	1141
Db	953	SKKPTFELT-----NGCYSPCSRS-----	971
Qy	1142	HTTNKLYPKDRVLDVSTRSPETKCPKONSIENDIEEKVSLASRGQPPTKSKTKGNDF	1201
Db	972	-----NPNRK-----YSP-----	980
Qy	1202	IDDSKLASADDIGTLICKNKKPLQIEESDITVSSSKSALHSSVPKSTNRDRTPLSRAM	1261
Db	981	-----MCRN-----	984
Qy	1262	FEGLKCDSESNSTLSESSDVTYSIODSSEEDMIVONSNEISIEQFRTREQDVEVLEPLK	1321
Db	985	-----	984
Qy	1322	ELVSGESTGNCEDRLPVKGTANGKKPSQOKLEERPYNKCSQDLKLTNTDKNNENRE	1381
Db	985	-----GYLWSAKQAHDKLEE-----	1001
Qy	1382	SEKKGORTSTFOJNGKDNKPKIYLKGECLKEISRVVSGVNPVKVNNINKIIPENDIKS	1441
Db	1002	-----	1001
Qy	1442	LTVKESAIRPFINGDVIMEDFNERNSETSKSHLLSSDAEBGNYRDSLETLPSTKESDQ	1501

Db 1002 ----- 1001
Qy 1502 TTPSASCPESNVQVEDMEIETSEYKVKVTSSPITSEESNLSNDFIDENGLPKNNEN 1561
Db 1002 -----SG 1003
Qy 1562 VNGESKRKVTITVTWTSTVATESKTVIKVEKDKQTVVSSSTENCAKSTVTTTTTVTK 1621
Db 1004 VLGEK----- 1009
Qy 1622 LSTPSTGSGVDIISVKEQSKTVVTTVTTDSLTGTGTLVTSMTVSKEYSTRDKVKLMKFS 1681
Db 1010 ----- 1009
Qy 1682 RPKKTSGTALPYRKYFVKTKSKSIFVLPNDLKLARKGGIREV--PYFNYNAKPAID 1739
Db 1010 -----AWPIPEIQTFTSKRGKSFVQLQKTLRQIMGGCQVQVMPGFSAGIKSNLL 1062
Qy 1740 IMPYPSRPFTFGITWRYLQTVKSLAGVSLMLRLWASLWDDMAAKVPPGGGSTRTETS 1799
Db 1063 IYYPAPRPPLDLCWKQTLNARSLHAVALQLKIWSIKFNEFD---PDTHPDRRVVI 1119
Qy 1800 ET--EITTTIIRKRDVGPIRGFYECIRKIIICPI-GVPEPKETPTPORKG-----LR 1850
Db 1120 DTPSHDERRILIRHKEMPPYQGYERYEMEIEIIPLYDEPEEDESLSRNRGSSSEFSHR 1179
Qy 1851 SSALRPKRPET-----PKQT-----GPVITWVAEELELWE----- 1883
Db 1180 SSSARKRPHOEFLSLKFGNPKKSKNAFRSLDNRRTAIRREWVDGVTLKVFEEKDYW 1239
Qy 1884 --IRAFARVEKEKAQV-----EQAKKRLQKQKPTVIATSTSPSTSTTSPAQ 1934
Db 1240 KWIRAEAKTAKRKLKLEATRAKAKADEERRIQQOQORSVARIPV-PMHS-----LIPSE 1294
Qy 1935 KVMVAPITSGSV---TTGKMWLTTKVGPATVTFQON-----KNEHGTFTATWKQOG 1983
Db 1295 RNNV-PYLSGQQORRNGNERGLEKYNSSSVSPAAGVASTPPGYHQPQNIIRQAG 1353
Qy 1984 SNGGVQVQOKVLGIIPSTGTSTQQTTFQPPRTATVTRPNTSGGGTTSNSQVITGPQ 2043
Db 1354 YNQ-----LPRKPTTSPFNQOS-RP-VATIPTTLPQLRAAGADG---VVRAVM 1396
Qy 2044 IRPG-----MTVIRTPLOOSTGLKAIIRTPWVOPGAPQVMTQIIRGO----- 2087
Db 1397 MTPGNKSTVNTSTPYQ-----ALNRQYQLQROQOQPAVRRLTNGYHFMGDMTMRGGGR 1451
Qy 2088 -----PVSTAVSAPNTVSTPGQKSLTSAT-----STSNLQSS 2120
Db 1452 NPSVQMHQRLPQNRALORPFGESTTEMRRVTEAIPDNDGDQPPVIRPYDPTSNFDAQ 1511
Qy 2121 ASQPPRQOGQVKLTMALQTLQTLQGHGNGQLFVVIQOGQTTGLQL-----IPQGVTVL 2176
Db 1512 RAQOOHPQSRPVYSTPAQMIITTPQGVKH--NVILMKASDGTQKMWLPKPGQFPPTGVI- 1568
Qy 2177 PGQGOQLQMANPNCVIVQRFELFTPLATTATTSTTTTSTTAAGTGEQKSL----- 2230
Db 1569 -STGQRVVPYRQPTAVQORQLYT-----ATPGTRVVRIPNANGAPRQODHOVMRRVQ 1621
Qy 2231 --SPQMVQHDK--TLPPAQ-----SSSVGPAKAQPTQAQPSARPQPTQPOS 2274
Db 1622 ASQPRAMEYDDGTPPGQOQRYVLOGNSGTPNVPKVSRRGPRGLTQMVOQQOQ 1681
Qy 2275 PAQEVQTOPEVOT---QTTVSSHVPSEAQ 2301
Db 1682 QHNPAYDMPDATTGFAVSTTTTQVPEEQ 1711

RESULT 10

Q95ZU8

ID Q95ZU8

AC Q95ZU8; PRELIMINARY; PRT; 1713 AA.

DT 01-DEC-2001 (T-EMBLrel. 19, Created)

DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE F26H11.2b protein.
GN F26H11.2 OR F26H11.2B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Barlow K.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RL Science 282:2012-2018(1998).
DR EMBL; Z61515; CAC42289.1; -.
DR WormPep; F26H11.2b; CE28009.
DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR004022; DDT_dom.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00384; AT_hook; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 1.
SQ SEQUENCE 1713 AA; 196824 MW; 178E470017D9AD5A CRC64;

Query Match 5.7%; Score 848.5; DB 5; Length 1713;

Best Local Similarity 17.4%; Pred. No. 5.8e-28;

Matches 434; Conservative 295; Mismatches 619; Indels 1147; Gaps 74;

Qy 4 EEEEE---DGAETQDSEDE---EDEMEEDDDSDYPMEEDD---DDASVCTE 52
Db 169 KRGEEDLIYDEDESEEESSDDFMLNDDQVVOEEELNLTVDIKELGDEENKYC-- 226
Qy 53 SSFSSHSTYSSTPGRKRPRVHRPRSPILER--KDIPPLEPKKSESDLMVNEHIMVIAI 110
Db 227 -----PWLDEDPASLPKLELPSSQDIPITASTMDAVEI 261
Qy 111 YEVLNFGTVLRUSPRFEDFCALVSEQOCTLMAEHVVLKAVLREEDTSNTTFGPAD 170
Db 262 YELRSYHRLTRITPTTFEDFCALISHNNSCIMAEVHMALLRNCLKSDDEEQTHYSVTE 321
Qy 171 LKDSVNSTLYFIDGMWPEVLRVYCSDKYHHVLPYQEA-----EDYPGPVENKI 222
Db 322 TNSVNMIMHMDTLTYAEILROYIEA-----YFADASVRDAINVDNYPFVGVDANI 374
Qy 223 K-----VLQFLVDQFLTNTIAREELMSEGVIOYDQHCVRCHK-LGDLIC 265
Db 375 QRDSEFFFNKHKVFLRVLLFMSYRFLYSSEFKKLVNNVKGKFNQDENCRCVCKSSGRVVG 434
Qy 266 CETCSAVYHLECVPKPLEEVPDEKQCEVCAHK-VPGVTDCAVETQKNKPYTRHPEIGY 324
Db 435 CTQCEAAHFVEC--SHLKPPPE-VLVCNICKKNSAVRGVLPPEAVDREP--LRSPIGR 489
Qy 325 DRSRRXYWFLNRLIIEEDTENENKIKWYSYTKVOLAEILCDLKDXYEAEKLCLEEM 384
Db 490 DRYGRYWFIVRLVYQSLDDETE---LYYSTVPQYQLQLKLDYKEDJCDTIRLR 545
Qy 385 REEIHHRMDITELTNKAGSNKSLAAANEELIESTRAKK---GQIDNVKSPETEKKOK 441
Db 546 IDEFLQMALTVEMTSERR-----EAALETWVKQLIGYDFAEATTQ----- 588
Qy 442 NETENDSKDAKNREEFEDQSLKSDDKTPD---DDEQKGESEETPEVGDKNGSVANL 498
Db 589 -----IYLHRDSMKRMASILRDCAQKGQKQEVKL----- 618
Qy 499 GDNVTTNATSEETSPSGRSPVGLSETPDSSNNAEKKVASELPQDYPPEEPNKTCSNNTS 558
Db 619 -----EE---PVGQSPVKCVQFVEDSI-----LPE----- 641

```
QY 559 ATTTSIQPNLENSNSSELNSQSESASAKAADDPENGRESHPTPVSIQEEIVGDPTSEKST 618
Db 642 -----SMIGIFADKLINTFWGGG-----TOEELVEQFV---DI 672
QY 619 GELSESPGAGKASGSTRIITRLRNPDSKLSQLKSQOAAAAHANKLFREGREVLVNS 678
Db 673 SDNFDAPSA-----681
QY 679 QGEISRLSTKKEVIMKGNINNYFKLQEGK---YRVTHNOYSTNSPALNKHQHRDHDKR 735
Db 682 -----NLWRMGDEGNDQTFMTYNYYSRNMSESEFLTRKKAADKK 721
QY 736 RHLAKHFCLTAPAGEFKW---NGSVHGSKVLTIITLRLTITOLENNIPSSGFHFENWASHR 791
Db 722 KIMASKF---AQIDNFDVVAKNRQFYGDASHLCKFIWTLQOVLKNIPIIDLHRRKWPEFA 779
QY 792 ANWIKAVQMSKPREFALALAILCAVPPVVMPLPIWREFLGHTRHLRMTSTIEREEKVKV 851
Db 780 KGFDELEVSVADDYKKLVTCLLKLDCAVRKTIFMPQWNGGLQTRLERITVDQRENFMKEQ 839
QY 852 KKEK-----QBEETHMQOATWVKYTFPVKHQVWQKGEYRVTG---YGGWSWLSKT 901
Db 840 QRLKKIDADALTOKLDDSFVRVNTYMKPKPNTY-ILRQGETYRNAGKSGMGWAWAAK 898
QY 902 HVYREVPLPGNTVNVYRKSLEGTNNMNDENDESDDRCSRPKIKIEPDSKDEYKG 961
Db 899 YVEKWI-----QVPESKPLPLAVTVEIKTESVSNRK---930
QY 962 SDAAGADONEMDISKITEKKQDVKELLDSDSKPCKEPEWEDDDMKTESHVNCDESS 1021
Db 931 -----ARLELLVSKIT-----942
QY 1022 QVDVVNVSEGHRLTSYKKTSSKLDGLLERRIKQFTLEEKQRLKIEGKIGKIGKT 1081
Db 943 -----KKQRSG-----GKS 952
QY 1082 STNSKNLSSEPVITKAREGQSDSMRQEQSPNANNDQPEDLIOGCSQSDSVLRMSDPS 1141
Db 953 SKKPFELT-----NGCYSPPSCRS-----971
QY 1142 HTTNKLYPKDRVLDDVSTRSETPCKPKONSIENDIEEKVSDLSRQGPPTKSKTGNDF 1201
Db 972 -----NPNRKC-----YSP-----980
QY 1202 IDDSKLAGADDIGTLICKNKKPLIOEESDTIVSSSKSALHSSVPKSTNDRDPTLSRAMD 1261
Db 981 -----MCRN-----984
QY 1262 FEGLGCDSESNSTLSESDTVSIQDSSEDMIVONSNEISIEOPRTREQDVEVLEPLKC 1321
Db 985 -----984
QY 1322 ELVSGESTGNCEDRLPVKGTANGKKPSQOKLEERPYNKCSQDLKLTNTDKKNENRE 1381
Db 985 -----GYLSAKOAHDERKLEE-----1001
QY 1382 SEKKGQRTSTFOINGKDNKPKIYLKGECLKEISRVVSGNVPEPKVNNINKIIPENDIKS 1441
Db 1002 -----1001
QY 1442 LTVKESAIRPFINGDVIMEDFNERNSSETKSHLLSSDAEGNRYRDSLETLPSTKESDSTQ 1501
Db 1002 -----1001
QY 1502 TTTPSASCPSNSVNOVEDMEIETSEVKVTYSSPITSBEESNLSDNFIDENGLPINKNEN 1561
Db 1002 -----SG 1003
QY 1562 VNGESKRKTIVTEVTMTSTVATESKTVIKVEKDGKQTVWSSTENCAKSTVTTTTTTVT 1621
Db 1004 VLGEK-----1009

1622 LSTPSTGSGVDIIISVKEQSKTVVTTTVDLSLTTTGGTLVTSMTVSKEYSTRDKVKLMKFS 1681
1010 -----1009
QY 1682 RPKTRSGTALPSYRKFVTKSTKKSIFVLPNDDLKLLARKGGIREV--PYENTNAKPAID 1739
1010 -----AMPPEIOTFTSTKRGKSIPLVLOKKILROMIMGGCCQVYMPGFSAGIKSNLL 1062
1740 IWPPSPRPRTGTIRYRLQTVKSLAGVSLMLSLMLASLWLDMAAKVPPGGGSTRTETS 1799
1063 IWYPAPRPLDLCWKQWTLNARSLHAVALQALAIWSSIKNEFD---PDDTHPDRRVVI 1119
1800 ET-EITTTETIIRRDVGPYGIREFYCIIRKIIQPI-GVETPKETTPQKRG-----LR 1950
1120 DTPSHDERRRIIRHKEMPPYQYERYEMEIEIIPLYDEPEEDESLSNRNGSGSEFSHR 1179
1851 SSALRKRPEP-----PKOT-----GPVIEETWAAEELELWE-----1883
1180 SSSARKKRQRHREFLSLKFGNYPKKSNAFRSLDNRATAIRREWVDGVTLYKFEIKDYW 1239
1884 --IRAFAEVEKEKAQAV-----EQAKKRLEQOQKPTVIATSTTSPTSSSTSTISPAQ 1934
1240 KWIRAEAEKTAKRLEATRAQAKADEERRRIQOQORSVARIPV-PMHS-----LIPSE 1294
1935 KVMVAPISGSV---TTGTMVLTTKVGSPATVTFQON-----KNFHQTFAATWVKQG 1983
1295 RNVV-PYLGSOQQRPNNGENERGLEKYNSSSVQAHGYASTPPPGYHQPQNIIRQAG 1353
1984 SNSGVVQVQKVLGIIPSSGTSTQOTFTSFQPTATVIRPNTSGSGGTTNSQVITGPQ 2043
1354 YNQ-----LPRKPTTSPFNQOS-RP-VATITPPLRAAGADG--VVRAVM 1396
2044 IRPG---MTVIRTPLOQSTLGLKAIIRTPVWVQAPQVMTQIIRGQ-----2087
1397 MTPCNKSTVNTNSTPYEQ-----ALNRQYQLOQOQOQPAVRRLTNGYHFMDETMRGGGR 1451
2088 -----PVSTAVSAPNTVSTPGQKSLTSAT-----STSNIQSS 2120
1452 NPSVQMHRQLPONRAALORPFGESTTEMRRVTEAAIPDNDGDEQPPVIRPYDPTISNDAQ 1511
2121 ASQPPRPQOQGVKLTMAQLTQLTQGHGNGOGLTVVIOGQGGTTLQQLQ-----IPQGVTVL 2176
1512 RAQOQHQPQSRPVYSTPAQIMRTITQPGVKH--NVILKASDGTQKMWLKPQGPPTGTVI- 1568
2177 PGPGQQLMAAMPNGTVQVORFLFTPLATATASTTTTTTSTVTAAGTGEORQSKL-----2230
1569 -STGQVVVYRQPTAVQOQRLYT-----ATPGTRVVRIPNANGGAPRQDHOVMRRVQ 1621
2231 -SPQ-MQVHODKTLPP-----AQSSSVGPAKAPQTAQPSARFQ-----PQTQ 2271
1622 ASGPRAMEYDDQGTTPPGQVRYVLOGNSGTENVAPPKVSSRSGRGLTMAQWVQOQ 1681
2272 PQSPAQ-----PEVQTQPEVOTQTVSSHVPEAQ 2301
1682 QHNFPQAHYMDPDDATGFAVSTTT---EQVDPDEQ 1713

RESULT 11
Q960Y3
ID Q960Y3 PRELIMINARY; PRT; 1022 AA.
AC Q960Y3;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE LD30146p.
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
```



```

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celisner S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Bazan J., An H., Baldwin D., Banon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreria S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Matel B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celisner S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003808; AAF58087.2; -.
DR FlyBase; FBgn0013988; Strn-Mlck.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR007110; Ig-Like.
DR InterPro; IPR003598; Ig-C2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 13.
DR SMART; SM00409; Ig; 21.
DR PROSITE; PS00835; Ig_Like; 20.
DR PROSITE; PS00290; Ig_MHC; 1.
KW Immunoglobulin domain.
SQ SEQUENCE 7210 AA; 808965 MW; 6B2E7395C07140D2 CRC64;

Query Match 4.3%; Score 641; DB 5; Length 7210;
Best Local Similarity 18.4%; Pred. No. 2,3e-18;
Matches 601; Conservative 538; Mismatches 1187; Indels 938; Gaps 135;

QY 5 EEEEDGDA-----EETQDSED-----DEEDMEEDDDSDYPPEMEDDDDDASVCT 51
D 52 ESFRSHSYSTPGRRKRVHRPRSPILEEK-DIPPLEFPKSSDLMPVNEHMVIAI 110
D 2619 KASAKQGGDIOEQSKQPEV-----SEVVAEKISEGKEIEPKKPEM----- 2661
QY 111 YVLRNFGTVLRSPFRFDFCAALYSQBCQTLMAEMHVVLLKAVLRREE--DTSNTTGP 168
D 2662 -----DTEAKSEKATVLDKQVLEKELEASAEKQGD 2692
QY 169 ADLKDSVNTLFIIDGTMTPVLRVYCESDKKEYHHVLPQAEADYPIYGVENKIRVQL 228

```

Db 3610 ---EAKPKKAKVLEKKSIEEAKLE-----DKKETQDTSADIDEKSQAEEVSE-TVSEKIT 3659
QY 1109 QEOSNANNDQPEDLTOGCSQSDSSVLRMSDPSTHTNKLYPKDRVLDVDSIRSPTKCPK 1168
Db 3660 DEKAQESQKEEVKD-----SEAKPK---KAKVLEKKSIEBEKLEDDK 3698
QY 1169 QNSIENDIEBK-----VSDLAS-----RQEPKTKSTKGNDFIDDSKLASADDITGLI 1217
Db 3699 EKOTESAIDEKSQAEEVSEIVSEKITDEKAQESQKKEVKDSEAKPKKAKVLEKKSIEEK 3758
QY 1218 CNKKPLIOBESDVIIV--SSKSALHSSVPKSTNDRDATPLSRAMDFEGKLGCDSPSNST 1275
Db 3759 LEDKK---ETQDTSADIDEKSQAEEVSEIVSEKITD-----EKAQESQKEEVKDSEAKPK 3809
QY 1276 LENSSTVIOQSDSEEDIMVONSISIQFRTREQDVEVLEPLKELYSGESTGNCEDR 1335
Db 3810 KAKVLEKKSIEBEKLEDDKKEKOTESAIDE-----KSQAEEVSE-IVSEKITDEKAQESQKK 3864
QY 1336 LPVKGTTEANGKPKSOQKLEERP--NKCSDQIKLNT---TDKKNENRESE-----1383
Db 3865 -EVKGEA---KPKKAKVLEKKSIEBEKLEDDK--KEKOTESAIDEKSQAEEVSEIVSEKIT 3919
QY 1384 -KKQRTSTFQNGKDNKPKIVLKGCLKEISESRVSVGNVEPKVNNINKIIPENDIKSL 1442
Db 3920 DEKAQESQKEEVKDSEAKPK---KAKVLEKKSIE-----BEKLENKEKOTESAIDEK 3969
QY 1443 TVKESAIRPFINGDVIIMEDFNERNSETKSHLLSSDAEGNYRDSLE--TLPSTKESDST 1500
Db 3970 SQAEEV-----SEIVSEKITDEKAQESQKKEVKDSEAKPKKAKVLEKKSIEBEKLEDDK 4023
QY 1501 QTTSPASCPESNSV-----NOVEDMEIETSEVKKVTSSPITSEENSLNDFIDENG 1553
Db 4024 EKOTESAIDEKSQAEEVSEIVSEINITDEKAQESQKKEVKDSEAKPKKAKVLEKKSIEBEK 4083
QY 1554 LPINKNENNGESKRKTIVETVMTSTVATESKTIVKVEKGDQKQVSVSTENCAKSTV- 1612
Db 4084 LEDKKTQDTSADIDEKSQAEEV-----SEIVSEKITDEKAQESQKKEVKDSEAKPKKAKVL 4139
QY 1613 ---TTTTTWTKLSTPGSDVJISVKQSKTWTVTVTDSLTITGGTLVTSMTVSKEY 1669
Db 4140 EKKSIEBEKLEDDKKEKOTESAIDEKSQAEEVSEIVSEINITDE-----KAQESQKK 4189
QY 1670 STRD-----KVKLM-----KFSRPKTSRGTALPSYRFVHTKSTKSFVLPNDLL 1715
Db 4190 EVKDSAKPKKAKVLEKKSIEBEKLEDDKKEKOTESAIDE-----KSQAEEVSEIVSEKI 4243
QY 1716 -KKLRKGIRPEVPYFNAYNAPALDIWPYPSPRPTGITWRYRLQVKSAGVSLMLRL 1774
Db 4244 TEEKAESQKKEVK--DSKAPK-----KAKVLEKKSIEEAKL-----4279
QY 1775 WASLRWDDMAKVPVPGGSTRTETSETITTEIIRKRDVPGYIRFEYCIIRKIIGPIGV 1834
Db 4280 -----EDKKET--QDTSADIDEKSQAEEV-----EIVSEKI-----4308
QY 1835 PETPKETPTPORKLSSALRKRPTPKQTGPVITETWVAEELELWEIRAFARVEKE 1894
Db 4309 --TDEKAQESQKEEVKDSEAKPKKAK-----VLEKKSIEBEKLE-----NKKKEQ 4351
QY 1895 KAAVEBQAKKRLQEQKPVIAVTSSTTSSTISPAQKVMVAPISGVTGTGKMWLT 1954
Db 4352 TESAIDEKSQK-----AEVSEIVSEKITDEKAQESQKKEVKS-----4389
QY 1955 TKVGSPTVTFQONKNFHTATWVKQGSNGVQVQKV--LGIIPSTGTSQOFTTSF 2013
Db 4390 --EAKPKKAKVLEKKSIEBEKLEDDKKEKOTESAIDEKSQAEEVSEIVSEKITDEK-----4442
QY 2014 QPRTATVTRPNTSGSGTTSNSQVITGPQIRPGWTVIRTPLOQSTLGHAIIRTPVMWOP 2073
Db 4443 -----AQESQKEEVKDSEAKPKKAKV---LEKKSIEBEKLEDDKKEKOT 4482
QY 2074 GAPPQVMTQIIRGPVPTAVSAPNTVSSTPGQKSLTSATSTNSIQSSASQPPRPOQ---2129
Db 4483 DSAIDEKSQ---KAEVSEIVS---EKITDEKAQES-----QKEEVKDSEAKPKKAKVLEKK 4532

QY 2130 --GOVKLTMAQLTOLTQGHGNGOGLTVVIOGOGQTTGOLQILPQGVTVLPQGOOLMOAA 2187
Db 4533 SIEEAKLEDDKKEKOTDS-----AIDEKSQAEEVSEIVSEKIT--DEKAQESQKEE 4580
QY 2188 MPNGTVQRELFETPLATATTASTTTTSTTAAGTGEOROSKLSQPMQVHQDTLPPAQS 2247
Db 4581 VKD-----SEAKPKKAKVLEKKSIEBEKLEDDKKEKOTESAIDEKSQAEEV 4625
QY 2248 SSVGPAKAPQOTAQPS-----ARPOPOTQPSAPQEVOTQPEVOTOT-----2290
Db 4626 SEIVSEKITDEKAQESQMEEVKDSEAKPKKAKVLEKKSIEEAKLEDDKKEKOTDSADIDEKS 4685
QY 2291 ---TVSSHVPSEAPQTHAOSKSPQVAAQSPQSNVQSQSPVRVQSPSQTRI---RSTPS 2344
Db 4686 QKAEVSEIVSEKITDEKAQESQKEEVKDSEAK-----PKKAKVLEKKSIEE 4732
QY 2345 QLSPGQSQSVOTT---TSQPIQPIHTSLQIPSQGQPSQ--POVQSTSTQTLSSGQTLNQV 2400
Db 4733 KLEDDKKEKOTESAIDEKSQAEEVSEIVSEKITDEKAQESQKKEVKGEAKPKKAKVLEKK 4792
QY 2401 SVSSPSRPOLOIQOPOPQVIAVPQLOQOVOLVSOIOSOVVAQIOAQS-----GVP 2451
Db 4793 SI---EELKLEDDKKEKOTESAIDEKSQAEEV-SEIVSEKITDEKAQESQKKEVKDSEAKP 4848
QY 2452 -----QKILQLPIQIOSSAV-----QTHQIQNVTVVQAASVQBLQARVQOQLRD 2496
Db 4849 KRAKVLKKSIEBEKLENKKEKOTESAIDEKSQAEEVSEIVSEKITDEKAQESQKKEVKD 4908
QY 2497 QOQKKQOQIEINVTNTPSKLLIKVEIIQOVVMKHNVAIEHLKQK-----KSTWPAE 2548
Db 4909 SEAKPKKAKV-----LEKKSIEBEKLEDDKKEKOTESAIDEKFKQAEVSEIVSEKITDEK 4962
QY 2549 REENORMIVCQNMVK---YILDK---IDKEEKQAARKREESVEQKRSK-----ON 2594
Db 4963 ARESKREEVKDSEAKPKKAKVLEKKSIEBEKLEDDKKEKOTESAIDEKSQAEEVSEIVSEK 5022
QY 2595 ATKLSALLPKHKBQLEAEILKKRALDLQDQIEVQBELKRDLEKKEKDLMLQALQATAVA 2654
Db 5023 ITDEKAQESQKKEVKDSEAKPKKAKILEKKSIEBK-----LDEKKEQK-----TETKVA 5072
QY 2655 APCPPVTVLPAPPAPPPPPPPGVOHTGLLSTPTLPVNASQKREKEEKDSSSKSKKK 2714
Db 5073 TDTKSTQ-----VEVSEIVLEKISIEEKAESQKVELKDSEAKSKKAK 5114
QY 2715 MISTTSKETKDTKLYCICKTPYDESKFYIGCDRCQNWYHGRVCVGLQSAEELIDEXVCP 2774
Db 5115 VLE-----KSTI-----LKEKLEDDKKEKQKEDGATN-----KSQKAEAD--VVP 5152
QY 2775 QCOSTEDAMTVLPTLTKDYEGLEKRLVRLSLQAHKNMWPFLPEVD-----PNDAPD 2824
Db 5153 EKISEEKVAEIKTP-----EPMDSKAKSKPQGLPADEKS 5186
QY 2825 YGVIVKEPMDLATMEERVORRYEKLTEFVADMTKIFDNCRYNPSDSFPYQCAEVLESF 2884
Db 5187 HGAKVSEVPVKNABEKTQLSAKKPTVLDEDLV-----PKRKP--YLAEQTADSI 5236
QY 2885 FVOKLKGf-----KASRSHNNK 2901
Db 5237 SLOTYKSMDSYKDRKESAKRKK 5260

RESULT 13

Q8MLD9

ID Q8MLD9 PRELIMINARY; PRF; 9270 AA.

AC Q8MLD9;

DT 01-OCT-2002 (T-EMBLrel. 22, Created)

DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE CG18255-PA.

GN STRN-MLCK OR CG18255.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

```

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan M.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q., Zhang L.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrieria S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnick S.E., Smith C.D.,
RA Tupy J.B., Bergman C., Bernier B., Carlson J.W., Celniker S.E.,
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn K., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RA EMBL; AE003808; AAM70936.1;
RA FlyBase; FBgn0013988; Strn-Mick.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; ig; 29.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00408; IGC2; 29.
DR SMART; SM00406; IGV; 4.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYRkc; 1.
DR PROSITE; PS00835; IG_LIKE; 26.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Immunoglobulin domain; Transferase.
SQ SEQUENCE 9270 AA; 1036749 MW; 0DD82D040B27C2AD CRC64;
Query Match 4.3%; Score 641; DB 5; Length 9270;
Best Local Similarity 18.4%; Pred. No. 3.le-18;
Matches 601; Conservative 538; Mismatches 1187; Indels 938; Gaps 135;
QY 5 EEEEDGDA-----ETQDSE-----DEEMEDDDSDSYPEMEDDDDDASYCT 51
DB 2568 ESQNQKDAAGDIRKSETEDVDVHSIEKKIEPKRSEKDKLDKDFLEEKE-----L 2618
QY 52 ESSFSHSTYSSTGRRKPRVHRPRSPITLEK-DIPLEFPKSSDEDVMPNHNIAI 110
DB 2619 KASAKKQGDQIEQSKQPEV-----SEVAAKISEGKIEEPKKPEEM----- 2661
QY 111 YEVLNFGTVLRLSPRFEDFCAALVSOBQCTLMAEMHVLLKAVLREE--DTSNTTFGP 168
DB 2662 -----DTEAKSEKATVLDKQVLEKELEASAERKQGD 2692
QY 169 ADLKDSVNSTLYFDGTWTPVRLRVYCESDKYHHVLPYQEAEDYPYGVENKIKVQLFL 228
DB 2693 QDVEKKSKQ-----PEVSEVAAEKISE-----ETIEPKKP----- 2723
QY 229 VDQFLTNIAREELMSEGVIQYDDHCRVCHLGLDLCCTCSAVYHLECKVPPLEEVPE 288
DB 2724 -----EVKTEIKSEKATALD-----KQVLEKELE 2749
QY 289 ---EMQCEVCVAHK--VPGVTDCVAE-----TQKNKPYIRHPEPIGVDRSRKRYWFLNR 336
DB 2750 ASAQKQCDQDVEKKSKQPEVSEIVAEEKISEKTIEEPKKPEVKTEIKSEKATA---LDK 2805
QY 337 RLIIETEDE-----NENEKKIWIYSTKVOLAELICLDKDYWEALCKILEEMREIHR-H 391
DB 2806 QVLBEKELEASAQKQGDQDVEKKSKQPEVSEV-----AE--KISEETIEEPKKPE 2854
QY 392 MDITEDLTNKGARNSKFLAAANEILLESIRAKKGDID----- 429
DB 2855 VKETEVEKSEKATVLDKQVL-----EKELEASAQKQGDQDVEKKFQKAEVSEVAAKISEET 2911
QY 430 -----NVKSPETEEDKNETENDSKDAEKNEEFEDQSLKDS----- 467
DB 2912 IEEPKKPEVKTEIKSEKATALDKQVLE--EKELEASAQKQGDQDVEKKSKQPEVSEVIA 2969

```


Db 549 AGATSASERSEBOECK -STEAPTSVDDIEPAKPTESSEASG-----EGEDVAKET 599
QY 578 NSQSBSAKAADDPENGERSHTPVSI-QEIV-GDFTSEKSTGELSESPGAKGASGT 635
Db 600 TPAGEASIAAGEEIVKGTTPAGEPSEGDEEIVKGTTPAESSESEDE----- 648
QY 636 RIITRLRNPDSKLSQLKSQOAAAAHEANKLEKEGKE--VLVNSOGEISRLSTKKEVIM 693
Db 649 --LTKVTTTTPAGESVAGEEIEAKETTPAGEPSIAGEEIVKVTTPAGE--SSIAGEEIVK 705
QY 694 KGINNYFKLQGBGKRVYHNOYSTNSFALNKHQHRHDKRRHLAHLKCLTPAGEFKWN 753
Db 706 -----VTTPAGESEGEIEIK-----VTTPAGESSSE 734
QY 754 GSVHGSKVIT-----ISTLRLTITOLENNIPSGFFHPNMAWRANWIKI--VOMCSKPRE 806
Db 735 G0EEIVKESTPAGEPISGEEDVIKATTSAPKSDIEGVKEPETATEVPAEEVEDFAKP-- 792
QY 807 FALALAILECAVKPVVMLPIWREFLGHTRLHRMTSIEREKEKVKKKKEQBEETMQOA 866
Db 793 ---TTPIAAEIEPIAGTPIPTDGTIS-----GEEIVKGTTPOTLEEQPEISEES 839
QY 867 TWKYVFPVKH0WK0GKEEYRTGYGGWISWKTHVYRFVPKPLGNTNWN-YRKSLEGT 925
Db 840 TEV----PVAED-----DLSSTASASIASSTEGV 865
QY 926 KNNMDENMDESKRKSRSPPKIKIPDSEKD-----EVRGSDAAKAGADQN 971
Db 866 QDAASETTSSAPARAGDKDEAATVPTAQDKDDEVEQDATDLPVEDVQSTAKTITTE 925
QY 972 EMDISKITEKKODVKELLDSDSKCEEPMEVDDMK----- 1010
Db 926 QPKEESTAEADAIEVTTSSPADKOEVPPEAPADKHDEEDVQATDLPKISDIGPPV 985
QY 1011 --TESHVNCQESQVDVWNVSEGFHL-RTSYKKTKSKLDGLLERRIKQFTL-----E 1061
Db 986 VTEATNGPETSDEATDKPSPVLPVPSQEVPSSTAKVNDNRNDFETEKTPLPSPGEDQ 1045
QY 1062 EKORLEKIKLEGGIK-----IGKTSNSSKNLSPIVITKAKEQCSDSMRQESPN 1115
Db 1046 SSEPLPAMDLPAGIPGEGCLVEGKTYANNITVPATAPCDVCK--CISSLVACQOMECK 1103
QY 1116 NND0PE-----DLIOGCSQS---DSSLVRLMSDPSHTTNKLPKDRVLDDVSIRSPETKC 1166
Db 1104 LPENLEKCTVAADLLDGCCTPYCDESTESAKEDEESTAK--PDNKIDEDVSEISTE-BI 1160
QY 1167 PKONSTENDIEEK-----VSLASRGQBPTKSKTKGNDFFIDDSKLASADDI 1213
Db 1161 PRDVIPTGITQPLSHVXPDEIOPVTVPQAQFDESTAKVDKBP--IDES----- 1210
QY 1214 GTLICKNKKPLIOESDITVSSSKSALHSVPKSTNDRDATPLSRAMDPEGLKGCDSSEN 1273
Db 1211 ----AEDKKPIGESEED-----SKPIDESEEDKKPVE----- 1238
QY 1274 STLENSDRTVSQDSSEDMIV-----QNNESISEQFRFREQD-----VEVLEPLKCE 1322
Db 1239 ---ESAEDKKPVEDSESEKPLFTVIPASEIEKESKEPEDEKTEADFAAPEEQEATTPA 1295
QY 1323 LVSGETGNCEDRLPVKGTANGK---KPSQOKKLEERPVNKCSDQI---KLKNTTDDKN 1376
Db 1296 QIADTAKEVDDKLATTSAPVGEDELKPADKERTKTE-----TAQIPDAEIPASTDEPE 1349
QY 1377 NENR-----ESEKKQRTST--FQINGKONKPKIYLKGECLKEISESRVSVGNVPKVNNI 1430
Db 1350 SSTELPTVLDKKPREDSKTGTEAPESDKVPEVPTSASTENEIEESDKFTTAVAPPKISAS 1409
QY 1431 NKLIIP--ENDIISLTVKE-----SAIRPFTNGDVI-----MEDFNERNSSETKSHL 1474
Db 1410 DETEPTAEDLVPATFPIESEFEVSTKKPVQGPPLTLAPAQPEKKPVDAETSTEADI 1469
QY 1475 LSSSDAEGNVR0SLFTLPSTKESDSQTTPP-SACPESNSNOVEDMEIETSEVKKKTS 1533
Db 1470 STEPSAEVEKEASGETSSEDNEIDAGSSTPVPVSADEDKTPTSTKTEVE---ADDKFTTV 1526

QY 1534 SPIT-SEESN---LSNDFIDENGLPI-----NKNENYNGESK-----RKT 1570
Db 1527 APLAGDEESNLPKLPQDIFEREA-PVAVTTAAPSDD---GEQKPVVEEKEPIEDGQKP 1582
QY 1571 VITEVTTMTSTVATESKTVIKKEGDKQI0VVSSTENCAKSTVTTTTTTVTKLSTPSTGGS 1630
Db 1583 IEDETSTPTS-----SENEIEPESDRATTIAPSKEE--PSEPSTGAPTDEPAEPSTADP 1635
QY 1631 VDIISVKEOSKTVTTTTVTDSLTGTLVTSMTVSKEYSTRDKVKLMKFSRPPKKTSGT 1690
Db 1636 ESDS-KEYPESEVPTTAPA---CEKPTSSITDEEPTATSPVAKPDEDEVEKETST 1690
QY 1691 ALPSYKFKVTSTKSKSIFVLPNDDLKKLARKGIREVPYFNNAKPAALDIMPYSPRPTF 1750
Db 1691 EIPTDAPASSEEDENS---STDQIPS-----EVP-----BKKP----- 1720
QY 1751 GITWRYLQTVKSLAGVSLMLRLMLWASLRWDDMAAKVPPGG---GSTRTETSETETITTE 1807
Db 1721 -----ETPAQTPEEGDIVGATAAPTTSDEVPVQ 1749
QY 1808 IIKRRDVGPIGRFYCIRKIIICPIGVPETPKETPTQPK-GLRSSALRPKRPTKOTG 1866
Db 1750 RLPE-----EVLAIPOPSTETGKIQODETTAAPSIDRK-- 1783
QY 1867 PVIIETWABEELELEWEIRAFARVEK--EKAQAVEQ-----QAKKRLQEQKPTVIATST 1919
Db 1784 ----EPYVEIDEATTVAPISEKDEKPTEEKPVQKPTGEBPSEEEKEKPIEDGVST 1839
QY 1920 TSPSTSTTTSIPAQKVMVAPISGVTGTMKVLVTTKVGSPATVTFQONKNFHQTATWV 1979
Db 1840 EGPVSTEAASEAGSTES-----SEEVKSTPEGEVAEKPED----- 1873
QY 1980 KQOSNS-GVVOVQKVLGIIPSTG---TSQOTFTSFOPRTATVIR-PNTSGSGGTTTS 2034
Db 1874 KOPSSTAQAPVETIPIESTELPAQDGDKPTSEAPVSDSDTSPSDEKIPSVSG----- 1927
QY 2035 NSQVITGPQIRPGMTVIRTPLQOSTLGKAIIRPVMVQGAPOQV-MTQIIRQOPVSTAV 2093
Db 1928 --EEVEGPEV-----ITASQAABEDELKTPAESPSSTDKVPETEYQKPEDETKAD 1977
QY 2094 SAPNTVSSTPGQKSLTSATST-----SNTQSSASQPPRQOGOVKLTMAQLTOLTQGHGG 2148
Db 1978 ETPESTVQSDVATSTAPVAGDIEKDEQATTASP--EEEEIKPTIAPAAEIPQ---- 2031
QY 2149 NQGLTVIIOGQGTTCOLOLIPQGVTVLP--GGQOLMO-AAMPNGTVORFLETPLATTA 2205
Db 2032 -----PSEKEPVDE-QEVESGTATPAESDQPIDELIATPATSPIDE-----ASTA 2076
QY 2206 TTASTTTTSTTAAGTGBQROSKLSPQM0VHODK-----TLPPAQSSSVGPAKAPQPT 2259
Db 2077 APTKEESTTVASAA-----SP--AVHDEIKDVTTPQVADKEVAAPQDETKT 2123
QY 2260 A-----QPSAR-PQPTQPOQSPAQEVOTOPVQF---QTTVSSHVPS-----EAQPTH- 2304
Db 2124 SIDVSTDSPTAQDQDEKQTEAPVAPTIVSSPTADSAASTSTVPEVPSVEIDTKPMD 2183
QY 2305 --AQSSKPOVA-AQSQPQSNVQOSPVRV--QSPSOTRIRPS-----TPSOLSPQOQS 2353
Db 2184 IMSQTTAPHTADGAASTSTDEDEQAPVTVPQDAEKTVPSPAPQSDSKTSPSEAPQDADE 2243
QY 2354 VQTTTSQ-----PIPIOPHTSLQIPSOQPOQSPQVQSSTQTL-----SSQTLNQVYS 2401
Db 2244 IPATATPLDNKNIPATVAPQTDGVPATAAPLDEDKIQTAAAPLDEEKIPSTAAAPLDDK 2303
QY 2402 VSSPSRQLOIQOPQPOQVIAVQLOQOVQVLSOIQSOVVAQIQAAQOSGVQKIKLQLPQ 2461
Db 2304 IPAPVSPVFDVEPSEKPAVSEYDGE-----ESTEPVHDVETSTDEPTSDAKLPPTS 2358
QY 2462 IQSSSAVQTHQIONVVTVQAA-----SVOELQRVQOOLRQOQKKQKQOIEINVNTPSKL 2516
Db 2359 APATPSESATAEIIVPETAAPLEKEVPEKATEQPELEKETPEKATEQPELEKETPEKA 2418

QY 2517 LIKVEI-----IQOVVVKHNAVIEHLKOKKSWTPAER-----EENORMIV 2557
 DB 2419 TEQPELEKETPEKATEQPELEKEVTDKATEQESVDEKTTPEPVVVKSLDSTEDE----- 2474
 QY 2558 CNQVMKYITLIDKKEEQAAKKRRESVEQKRKQKQKATLKSAL----- 2601
 DB 2475 -----ESVESEESADKKKDKKETEEDTDKKHEPEVPVAVVSEIPOPSEEAAPTIG 2525
 QY 2602 --LFFKHQDLRAELIKKRALDKDLQIETOVELKRDILKIKKEDKMLQALQATAVAAACPP 2659
 DB 2526 HPLFPH-----LASSTTPPAVDR-----VGEE-----DEENTVTKLSSTTTTSTTESP 2570
 QY 2660 VT-----PVLPAAPPAPPPPPGQVHGGLLSTPLTPVASKRKRREKDSKSSKKKKM 2715
 DB 2571 VTSAPSTTVASQQOQPTTPPYG--HA-----PEYEDYDEEEVFGFCRCRYAGKL 2620
 QY 2716 ISTTSKTKDKTKLYCKTCKPYDESKFYIGDCRCQNVYHRCVGILOS-----PAEL 2767
 DB 2621 VYSAQIIPRDDPCDFCF-----FRSDIICLQOQSCPPPIAGCHEEP 2661
 QY 2768 IDEYVCPOC-----STEDAMTVLTP 2788
 DB 2662 ISGFCPCRYECPVSMVAALNITTTTSTTTPP 2695

RESULT 15
 Q9NJ17 PRELIMINARY; PRT; 5476 AA.
 AC Q9NJ17;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Split ends (CG18497-PC).
 GN SPEN OR CG18497.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20157049; PubMed=10655223;
 RA Rebay I., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Laverty T.,
 RA Suh C., Voss M., Williams A., Rubin G.M.;
 RA "A Genetic Screen for Novel Components of the Ras/Mitogen-Activated
 RT Protein Kinase Signaling Pathway That Interact with the Yan Gene of
 RT Drosophila Identifies split ends, a New RNA Recognition Motif-
 RT Containing Protein";
 RL Genetics 154:695-712(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., M.G.,
 RA Palazzolo M., Pittman K.S., Pan S., Pollard J., Puri V., Reese H.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J., Ye J.,
 RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong W., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez C., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragass V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M.B., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF184612; AAF26299.1; -;
 DR EMBL; AE003590; AAN10511.1; -;
 DR HSP; P09851; 1HA1
 DR FlyBase; FBgn0016977; spen.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 3.
 DR SMART; SM00360; RRM; 3.
 DR PROSITE; PS0102; RRM; 3.
 SQ SEQUENCE 5476 AA; 590531 MW; 93FAA8C7860770C2 CRC64;

Query Match 4.1%; Score 613.5; DB 5; Length 5476;
 Best Local Similarity 19.5%; Pred. No. 2.4e-17;
 Matches 546; Conservative 405; Mismatches 1013; Indels 841; Gaps 118;
 QY 305 DCVAEIQNKPYIRHEPIGYDRSRPKYWFLENR-----RLIIEEDT 344
 DB 1855 TADAKAKNQ-----RHEKKKKQKREKDKLRKQVEREKDKAKQOEKEKDKAKKEEK 1911
 QY 345 ENENEKKIWIYSTKVQLAEILDCLDKDYWEAPLCKILEEMREITHRHMDITE-DLTNKA 403
 DB 1855 TADAKAKNQ-----RHEKKKKQKREKDKLRKQVEREKDKAKQOEKEKDKAKKEEK 1911
 QY 345 ENENEKKIWIYSTKVQLAEILDCLDKDYWEAPLCKILEEMREITHRHMDITE-DLTNKA 403
 DB 1855 TADAKAKNQ-----RHEKKKKQKREKDKLRKQVEREKDKAKQOEKEKDKAKKEEK 1911

Db 1912 EREREKKAQDREK-----KEREREELRE--KEORDKEQKEIREKDLREKEQ 1958
QY 404 GSNKSFIAAEEILESIRAKGDIIDNVKSPETEEDKNETENDSKDAKNREDFEQSL 463
Db 1959 -----BERONREKELDKOLUREKEMEREKEQREKELHREKDOREHR 2000
QY 464 EKXSDDKTPDDOPE--QGSEEPTE-----VGDKGNSVANLGDNTNATSETSPS 513
Db 2001 EKQSRAMDVBQEGRGMRRELSSYQKSKMDIAGEASSLTAIDCQHKNENAMDTIAOQT 2060
QY 514 EGRSPVGCLESEYPDSSNMAEKKVASLPODVPEPNK---TCBSSNTSA-----TTTSI 564
Db 2061 PGASP-----STP--SDNTPKRSRKLRSNPVRLHRLRLSSQESNHSAGGSCGSSH 2113
QY 565 QPNLENSSELNSQSESAKAADPENGERS--HTPVSIQOEIVGFTSEKSTGELS 622
Db 2114 QIHHEDEVKRIRMENSONISVSHSNQRLNDRRDSKEHKSSFKED-----KNSSSHIS 2166
QY 623 ESPGACKGASGSTRITRLRNPDSKLSQLKSQOAAAAHEANKLFEGKEVLVNSQGEI 682
Db 2167 RPHCGGSSASSKHHHR-----RDKHHQKGSASSIETNSSI-----EVVVDPISTQTK 2214
QY 683 SRLSTKKEVIMKGINNYEKLQEGKYRVYVHNOYSTNSFALNK-----HOHR----- 729
Db 2215 HNLNTESELOSHQ-----PAREKEREHFSSHANSSSRHKSXKRDHHRHREKKRHSVA 2267
QY 730 -----EDHKRRHLRAHKFCLT-----PAGEFKWNGSVHGSKVITISTRLTITQLENNIP 779
Db 2268 ESTINTDEHTPOQHNPHRIRISAAGSGSAGELSSAANTSSGKLHQHHRSSVERKSSRGS 2327
QY 780 SFFPHNWAHRANWIKAVOMGSKPREFALAILALECAVKPVVMLPIWREFLGHTLRLHM 839
Db 2328 DEGHSSSKSLAKLMLLSADSDDDTDDA-----SKKHSI 2362
QY 840 TSIERE-----EKEKVKKKEKQEBEETMQOATWVKYTPPVKHVWKQGEYRVTVGYG 894
Db 2363 FDIPODCPNVMYDKVAKSCNMQOAEKKIAKFS-----QLKQRAKKRSTSYDG 2417
QY 895 WS---WISKTHVYRFVKPLPGNTNVNYKSLGCTKNMNDENMDESDKRCSRPKKIKIE 951
Db 2418 DSDTEPEDRQH-----RNSGSSSFHGR--YPLGSSDDDDDETHORRIS----- 2460
QY 952 PDSERDEVKSDAAKADQNE-----MDISKTEKKDQDVKELLSDSDSKPCKEPEVMD 1006
Db 2461 -----SDSDAHEGGQDNQCASTLADANRVQMO--QNLRLRCGDGDS--SED--EIR 2506
QY 1007 DMKTESHVNCQESSQDVVNVNVEG-----FHLRTSY-----KKTSSKLDGLL 1051
Db 2507 RVNMKSHFGKRNSTRIASSESOSQAPDLTIKQEHPIAPAQEIKEQLSDEQKFK 2566
QY 1052 ERRIKOTLEEKORLEKIKLEGGIKIGIKTSTNSSKNLSESPVITKAKEGCSQSDSMRQEQ 1111
Db 2567 SRHDSNSSIEER-----KLATEREI-----NTELGDFYNSSEYTVYTKLKEYSPETRKKHKK 2618
QY 1112 SPANNDQPEDLIQGSQSDSS-----VLRMSDPS-----HTNKLKYPKDRVLDVDSI-- 1159
Db 2619 SKRR-----LKSSSTADTSAQAQPLVMTPLTPSIFDVHSSECKTRFDNFDDLKTEC 2670
QY 1160 -----RSPETKCPKONSTENDIEEKVSDLASRGOEPTKSKTKGNDFDIDDSK 1206
Db 2671 SSIPLEISAGERKKHKEKREKRLNRNTEATVPN-----SPTNNDT-----SSE 2716
QY 1207 LASADDIGTLICKNKKPLQIESDITIYSSSKSALHSSVPKSTNDRDATPLSR----- 1258
Db 2717 KLSKEERHLKSKSKSDMNSCNTKIYNSGA-H---PSTSPSLPATPTSAPSTAQSK 2772
QY 1259 ----AMDFEGKLCDSSESNTLNSSDTVSIODSEEDMIVONSNESISEQFRTDRQVE 1314
Db 2773 RGEDKMEFIFGIISDEESQFPEQAETNKDIIPSS-----VSTTGPVISAALQTYKQ--- 2824
QY 1315 VLEPLKCLVSGESTGNCEDRLPVKCTEANGK-----KPSOOKKLEERPYNKCSQDLKLN 1370
Db 2825 --EP-----STPNS-----KNEEAHQITVHEPEQOQOOLER---SRLSGGSSSS 2864
QY 1371 TTDKKNENRESEK-----KGQRTSTFTQINGKDNKPKIYILKGECLEISESRVVVGNVEPK 1426
Db 2865 HADREHRRREKREKKREKRSQREQQNOIQHOKSK-----VETK 2902
QY 1427 VNNIKIIPENDIKSLTVKESAIRPFIINGDVIEMEDFNERNSSSETKSHLLSSSDAEGNYRD 1486
Db 2903 VDDNNSVDMDACRALEAQ-----LMSDFDTKPISE----- 2933
QY 1487 SLETLPSK---ESDSTQTTTPSASCPESNSVNOV-----EDMEIETSEVKVTSPPIT 1537
Db 2934 --EATSTAAITYSDMTDVFRESDN--EDNNSVDWTKQGVKSEQEQHKKDKKKKRSK 2990
QY 1538 SEESNLNDFIDENGLPINKNENNGESKRKTVITEVTMTSTVATESKTVIKVKGDK 1597
Db 2991 EEKQEKLL-----QOORRESLPNVASTSSAPPTPGKLTNVVQAASK 3031
QY 1598 QT-----VVSSTENCAKSTV-----TTTTTTVTKLSTPGSGSDVILISVKEQS 1640
Db 3032 HADQLDAKHIISSPPVCKPSPSLPOLIGDDDDDALHTPKAKPTTPSSRGNDGLTPSREKP 3091
QY 1641 KTV-----VTTTVDLSLT-----TTGTLTVTS----- 1662
Db 3092 RLISPIPKPTPIANSSTLSTQSAETPVSSGTVISSSALATTPTSSTAAGVSAAPGLDNP 3151
QY 1663 -----MTVSKEYST-----RDKVKLMKFRPKK 1685
Db 3152 TSASAOCKKESFIPGFDGQLDDRISAVOSISAEFNSTSLDNDIADBEKIPVASPPRA 3211
QY 1686 TRSGTALPSYKFEVTKSTK----- 1707
Db 3212 TKPLDLEESKSVTISQEEETSAVSALLGESFTSTTDDYSLDGDMESSVNELETPTL 3271
QY 1708 FVLNPDDLKKLARK-----GGIREVPVENYNAKPALDIWVPSP-----RPTFGIT 1753
Db 3272 VTAEPDEEALAAKALETAGEPASILEEP-----EMEPERAEAPDPOPEAIESEPVEVL 3327
QY 1754 WYRL--QTVKSLAGVSLMLRLWLASLRWDDMAAKVPPGGSTRTSETSEITTEITIKRR 1812
Db 3328 DPEELNKAQVSLKHEDMM-----DIKADTPQSERDLQIDT--DTEENPDE---AD 3372
QY 1813 DVGYPGIREYCIKRIICIPGVPETPKETPTPQKGLRSLALRPKRPETPKQGPVLIET 1872
Db 3373 SSGP-----SLKIDETVQSSSSPEKSISSNS-----PTRETANIDI-- 3409
QY 1873 WVAEELELWEIRAFARVEKEKAQAVEQOAKKRLQOKP-----TVIATSTTSPTSST 1926
Db 3410 -----PNVESQPKLSNESTPQPSVITKLPFLDTPKTVPAGLPPSPVKIE 3453
QY 1927 TSTISPAQVMVAPISGSV-----TTGTMVLTTKVGSPPATVTPQQKNKPHQTATVWKQG 1982
Db 3454 PPTISKLOQLVQPVQTVLPAPHSTGSGISANSVINLDSNVISSCSNTSAASATASASA 3513
QY 1983 QNSGVVQVQOQVILGIIIPSTGTSQOTFTSFQOP--RTATVTIRNPTSGSGGTTSN---SQV 2038
Db 3514 SISFGSPTASONAM---PQASTPKGPITPQQAIRTQSLIMOPPTISPEQTFHFVAPQM 3570
QY 2039 ITGPQI---RPG--MTVIRTPLOQSTL---GKAIIRT-----PVMVQPGAPQ 2078
Db 3571 VLSPOSHHPQPGQTYMVGIRAPSPHLSHPGSGVAQSLVGLQSLVGRPMWSQP--SPQ 3629
QY 2079 VMTQIIRGQPVSTAVSAPNTVSTPGOKSLTSATSTSNIOSSA---SQPPRPOQGVKLT 2135
Db 3630 ---QVQOTQOQHAIITSPQSSNISPLASPTTRVLLSSNSPTTSKVNSYQPRNQ--- 3680
QY 2136 MAQLTOLTQOCHGNQGLTVVIOCGQGTTLQOLQILPQGVTVLPQCGQOLQMAAMPNGTVQR 2195
Db 3681 -----VPQPSKPSVAEVQTPQLMTI---PLOKMTPIQVPHPT-- 3717
QY 2196 FLFTPLATTATTATTTTAACTGEGORQSKLSPQMVHQDKTLPAPQSSSVGPAPA 2255
Db 3718 -IISKVTVVQPOQATOSQVASSPPLGS-----LPPHKNVH-----LNAHQN-----QQ 3759

FT	ZN_FING	251	298	PHD-TYPE.	
FT	DOMAIN	4	9	POLY-GLU.	
FT	DOMAIN	30	35	POLY-ASP.	
FT	DOMAIN	42	46	POLY-ASP.	
FT	DOMAIN	657	660	POLY-ALA.	
SQ	SEQUENCE	810 AA;	91799 MW;	F7E2C992FE5BE96D CRC64;	
Query Match					27.3%; Score 4082; DB 1; Length 810;
Best Local Similarity					98.7%; Pred. No. 7e-137;
Matches 7/77; Conservative 4;					Mismatches 2; Indels 4; Gaps 1;
QY	1	MYSEEEEDGDAETQSDDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHST	60		
DB	1	MYSEEEEDGDAETQSDDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHST	60		
QY	61	YSTGGRKPRVHRPRSPILEKDIPLPEFFPKSSDLMVPHNEIMNVIAIYEVLENFTV	120		
DB	61	YSTGGRKPRVHRPRSPILEKDIPLPEFFPKSSDLMVPHNEIMNVIAIYEVLENFTV	120		
QY	121	LRLSPRFDFCAALVSQOCTLMAEMHVLLKAVLREEDTNTTFFGADLKDSVNSTLY	180		
DB	121	LRLSPRFDFCAALVSQOCTLMAEMHVLLKAVLREEDTNTTFFGADLKDSVNSTLY	180		
QY	181	FIDGMTWPEVLRVYCESDKKEYHHVLPYQAEADYPYGPVENKTKVLFQFLVDQFLTTNIARE	240		
DB	181	FIDGMTWPEVLRVYCESDKKEYHHVLPYQAEADYPYGPVENKTKVLFQFLVDQFLTTNIARE	240		
QY	241	ELMSEGVIOYDHCRCVCHKGLDLCCTCSAYVHLECVKPPLEEVPEDEWQCEVCAHKV	300		
DB	241	ELMSEGVIOYDHCRCVCHKGLDLCCTCSAYVHLECVKPPLEEVPEDEWQCEVCAHKV	300		
QY	301	PGVTDCAVETQKNPKYIRHEPIGYDRSRKRYWFLNRRLLIEEDTENENEKKIWIYSTKVQ	360		
DB	301	PGVTDCAVETQKNPKYIRHEPIGYDRSRKRYWFLNRRLLIEEDTENENEKKIWIYSTKVQ	360		
QY	361	LABELIDCKDYWEALCKILEMREIHRHMDITEDLTNKGSKNSFLAANEEILLES	420		
DB	361	LABELIDCKDYWEALCKILEMREIHRHMDITEDLTNKGSKNSFLAANEEILLES	420		
QY	421	IRAKKGDIDNVKSPETEKKNETENDSKDAKNREEFQDSLEKSDSDKTPDDPDEQCK	480		
DB	421	IRAKKGDIDNVKSPETEKKNETENDSKDAKNREEFQDSLEKSDSDKTPDDPDEQCK	480		
QY	481	SEEPTEVGKGNVSANLGDNTNATSEETSPSEGRSPVGCISLSETPDSSNMAEKKVASEL	540		
DB	481	SEEPTEVGKGNVSANLGDNTNATSEETSPSEGRSPVGCISLSETPDSSNMAEKKVASEL	540		
QY	541	PDVPEEPNKTCESSNTSATTTISIQPNLENSSSSELNSSQESAKAADDPENGERSHT	600		
DB	541	PDVPEEPNKTCESSNTSATTTISIQPNLENSSSSELNSSQESAKAADDPENGERSHT	600		
QY	601	PVSIQIEIVGDFTSKSTGSELSGAGKSGASTRIITRLRNPDSKLSQKSSQVAAAA	660		
DB	601	PVSIQIEIVGDFTSKSTGSELSGAGKSGASTRIITRLRNPDSKLSQKSSQVAAAA	660		
QY	661	HEANKLFKEGKEVLVNVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKRYVHNOYSTNS	720		
DB	661	HEANKLFKEGKEVLVNVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKRYVHNOYSTNS	720		
QY	721	FALNKKHQHREDHDKRRLAHKFCLTTPAGEFKWNGSVHGSKVLTITSLRITITOLENNIPS	780		
DB	721	FALNKKHQHREDHDKRRLAHKFCLTTPAGEFKWNGSVHGSKVLTITSLRITITOLENNIPS	780		
QY	781	SFFHPNWN 787			
DB	777	TSLHPSF 783			
RESULT 2					
TRX_DROME					
ID	TRX_DROME	STANDARD;	PRT;	3726 AA.	
AC	P20659;	Q27255;	Q27327;		
DT	01-FEB-1991 (Rel. 17, Created)				

DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	28-FEB-2003 (Rel. 41, Last annotation update)	
DE	Trithorax protein.	
GN	TRX.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	MEDLINE=9019757; PubMed=2107543;	
RA	Mazo A.M., Huang D.-H., Mozer B.A., Dawid I.B.;	
RT	"The trithorax gene, a trans-acting regulator of the bithorax complex	
RT	in Drosophila, encodes a protein with zinc-binding domains.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 87:2112-2116(1990).	
RN	[2]	
RP	SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.	
RA	MEDLINE=95009521; PubMed=7924996;	
RA	Sedkov Y., Tillib S., Mizrokhi L., Mazo A.;	
RT	"The bithorax complex is regulated by trithorax earlier during	
RT	Drosophila embryogenesis than is the Antennapedia complex, correlating	
RT	with a bithorax-like expression pattern of distinct early trithorax	
RT	transcripts".;	
RL	Development 120:1907-1917(1994).	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RA	STRAIN=Oregon-R;	
RA	MEDLINE=96100387; PubMed=8555104;	
RA	Tillib S., Sedkov Y., Mizrokhi L., Mazo A.;	
RT	"Conservation of structure and expression of the trithorax gene	
RT	between Drosophila virilis and Drosophila melanogaster.";	
RL	Mech. Dev. 53:113-122(1995).	
RN	[4]	
RP	CHARACTERIZATION.	
RA	MEDLINE=95047388; PubMed=7958911;	
RA	Kuzin B., Tillib S., Sedkov Y., Mizrokhi L., Mazo A.;	
RT	"The Drosophila trithorax gene encodes a chromosomal protein and	
RT	directly regulates the region-specific homeotic gene fork head.";	
RL	Genes Dev. 8:2478-2490(1994).	
CC	-!- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION	
CC	WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.	
CC	IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.	
CC	-!- SUBCELLULAR LOCATION: Nuclear.	
CC	-!- ALTERNATIVE PRODUCTS:	
CC	Event-Alternative splicing; Named isoforms=2;	
CC	Name=Long;	
CC	IsoId=P20659-1; Sequence=Displayed;	
CC	Name=Short;	
CC	IsoId=P20659-2; Sequence=VSP_006665;	
CC	-!- MISCELLANEOUS: THIS PROTEIN HAS BEEN EXPERIMENTALLY SHOWN TO BIND	
CC	ZINC.	
CC	-!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.	
CC	-!- SIMILARITY: Contains 1 SET domain.	
CC	-!- SIMILARITY: Contains 5 PHD-type zinc fingers.	
CC	-!- SIMILARITY: Contains 1 post-SET domain.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch)	
CC	-----	
DR	EMBL; M31617; AAA29025.1; -	
DR	EMBL; Z50152; CAA90514.1; -	
DR	EMBL; Z50152; CAA90513.1; -	
DR	EMBL; Z31725; CAA83516.1; -	
DR	EMBL; Z31725; CAA83515.1; -	
DR	PIR; A35085; A35085.	
DR	HSSP; P20393; IA6Y.	
DR	TRANSFAC; T00850; -	

DR FlyBase; FBgn0003862; trx.
DR InterPro; IPR003889; Fyrich_C.
DR InterPro; IPR003888; Fyrich_N.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR001214; SET.
DR InterPro; IPR001965; znf_PHD.
DR Pfam; PF00628; PHD; 3.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00542; FYRIC; 1.
DR SMART; SM00541; FYRN; 1.
DR SMART; SM00249; PHD; 5.
DR SMART; SM00508; PostSET; 2.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS00868; POST_SET; 1.
DR PROSITE; PS0280; SET; 1.
DR PROSITE; PS01359; ZF_PHD_1; 4.
DR PROSITE; PS00016; ZF_PHD_2; 3.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Developmental protein; Activator;
KW Alternative splicing.
FT ZN_FING 1266 1347 PHD-TYPE 1.
FT ZN_FING 1348 1393 PHD-TYPE 2.
FT ZN_FING 1421 1482 PHD-TYPE 3.
FT ZN_FING 1734 1793 PHD-TYPE 4 (ATYPICAL).
FT ZN_FING 1794 1844 PHD-TYPE 5 (ATYPICAL).
FT DOMAIN 3587 3708 SET.
FT DOMAIN 3710 3726 POST-SET.
FT DOMAIN 512 516 POLY-SER.
FT DOMAIN 565 570 POLY-ASP.
FT DOMAIN 661 664 POLY-SER.
FT DOMAIN 901 910 POLY-SER.
FT DOMAIN 1576 1582 POLY-GLN.
FT DOMAIN 2298 3027 GLN-RICH.
FT DOMAIN 3032 3040 POLY-SER.
FT DOMAIN 3181 3184 POLY-GLN.
FT DOMAIN 3220 3225 POLY-GLU.
FT VARSPIC 1 368 Missing (in isoform Short).
FT CONFLICT 2025 2025 P -> PWLTSPKFLGLSTHGGLLWLLGLLVVRLKQG
FT CONFLICT 2341 2341 R -> S (IN REF. 1).
FT CONFLICT 2392 2392 G -> S (IN REF. 1).
SQ SEQUENCE 3726 AA; 400575 MW; D2756E50763D1CF5 CRC64;

Query Match 3.2%; Score 476; DB 1; Length 3726;
Best Local Similarity 18.8%; Pred. No. 1.9e-09;
Matches 628; Conservative 419; Mismatches 1138; Indels 1154; Gaps 155;

QY 5 EEEEEGDAAETQDSDD-----EEDMEEDDDSDYPEMEDDDDDASCTESSFRSHST 60
DB 538 DEDEGGVTFRNDSPEDQNNAEDEDDDEAEEDDDQNEEDNDDEAAESEKSAETKS 597
QY 61 YSS-----TPGR--RKPRVHRSPRIIEEKDIPPLEFPKSSIDLMPN 101
DB 598 AGADERDPEKQLVMDSHFVLPKRSTRSSRIIKPNKRILLEGA-----STKKPLSLGD 651
QY 102 EHMNVIAI-----YEVLNFGT-----VLRSPFRFE 129
DB 652 SKGKNVFGTSSSSAGSTASTFSASTNLKLGKTEFFNFGTLKPNSSAAGIFVLRQPRLQFQ 711
QY 130 DFCALVSOEQTLMAEHVVLLKAVLREEDTSNTFFGPADLKDSVNSTLYIFDGMTWPE 189
DB 712 -----ADNQATFAAPKACPTSPSAIPKPAISLATSSFGSLASTNSSTV-----TPTS 760
QY 190 VLRVYCE---SDKEYHVLVPEAEADYPYGPVENKI--KVLQFLVDQFLTTNIAREELMS 244
DB 761 ACST-CAVVSKEV-----TQARKYGVVACDVCRAFFKFKTKKSISANSS-TAN-TS 810
QY 245 EGVYQYDHRCHVCHKIGDLLC-----CETCSAVYHLECVKPPLEVEPDEWQCEVC-- 295
DB 811 SGSQY-----LQCKGNEGSPCSIHSKSLQKNFKFYKD-----RCTACWL 852
QY 296 -----VAHK-----VPGVT-DCVAEIQKNPKYIRHEPIGYDRSRKRYWFLNRR 337

DB 853 KKMISFQLPAAHRSRLSAILPGMRGEAAAREEKSALL--SPTGSLRF----- 900
QY 338 LIIEEDTENENKKIWYSTKVLQELADCLDKDYWEA--ELCKILIEEMREETHRHMDIT 395
DB 901 ----TSTASSSPSV-VASTSVK-----WKSGDSTSAITSIKPNPLAENNVIT 943
QY 396 EDLTNKARGS---NKSFL-----AAANEELIESIRAK--KGDIDNVKSPETEKK 441
DB 944 FGSTPLLRPAILENPLFLKISNAADQKLAARAEISPLTKKNSKQEKVKESEQSEKLL 1003
QY 442 NETENDSKD---AEKNREEFEDQSLEKSDDDKTTPDDDPQGGKSE-----EPTVEGD 489
DB 1004 SPTQAGTKKSGAAEAQVEEVQPKKEAPQTSITTPTPSASNGASHGVPQAEAGETNATGD 1063
QY 490 KGNVSANLGDNTTATSEETSPSEGRSPVGCLE--TPDSSNMAEKKVASELPODV--- 544
DB 1064 TLKQRIDLPGRVKHVCRSASIVLQG-PLATFGEQOQDEADADMQEETAAAPVPSAIMEP 1122
QY 545 -PEEPNKTCESSNTSATTTSIQPNLENSNSSELSSQESAKAADDPENGERSHPTVS 603
DB 1123 SPEKPTHIIVTDENDNCASCKTSPVGDESKPSKSSGSAQAQAEVKA-----TA 1168
QY 604 IQEEIVGDFTSEKSTGELSESPGAGKAGCSGSTRITRLRNPDSKLSQLKSQQVAAAHEA 663
DB 1169 LGKE-----GTASAAGSGSAKVT-TRN-----AASAS 1194
QY 664 NKLFBKGKVLVNVNDSOGEISRLSKKEKIMGNINNYFKLGOEGKYRVYHNOYSTNSFAL 723
DB 1195 NLIVAASKK---QRNGDI---ATSSSVTQSSN-----QTQGRKTEKHQORT----L 1236
QY 724 NKQHREHD-----KRRHLA-----HKFCLT 745
DB 1237 ISIDFENYDPAEVCOTGFLIVETVAQALCFLCGSTGLDPLIFCAACCEPYHQYCVQ 1296
QY 746 PAGEFKWNSVHGS---KVLTTTLTLTITOLENNIPSSFFFPNASHANWI----- 795
DB 1297 DEYNLK-----HGSEFDTTLMGSLLETTVN--ASTGPPSSL--NQLTQRLNMLCPRCTVC 1347
QY 796 -----KAVQMCCK-----PREFALALALECAVRPVVMLPIW 827
DB 1348 YTCNMSSGSKVKQCKQKNVHSTCLGTSKRLGADRPCLVCNCLCKCSSTKV----- 1401
QY 828 REFLG-----HTRL----- 836
DB 1402 SKFVGNLPMCTGCFKLKRGKGNFCPCQRCYDNDNDFDKMMEGDCQGVHSCGLSDEQ 1461
QY 837 HRMTSIEREEKEKVKKKEKQEEETMQOATWVKYTFPVKHQVWKQK-GEERYVTGCGW 895
DB 1462 YNLLSTLPESIEFICKKCARNESS-----KIAEEMQAVMEEFKASLYSVL 1509
QY 896 SWISTKHVYRFPKLPNGTNNVYKSLGCTKNNDENMDSDKRCRSRSPKKIKIEP--- 952
DB 1510 KLLSKRQACALLKLSPRKNV-----RCTCGASNQKQLQPKAL 1548
QY 953 --DSEKDEYKSDAAKADQNDMDISKITEKDDQVKELLDSDSKPCKEPEMEVDDMK 1010
DB 1549 QFSSGSDNGLGSD---GESONSDVVEFKDQOQO---QOQRNANMNKP-RVKPL-----PCS 1598
QY 1011 TSHVNCQRS-SQVDVNVNVEGFLHRTSYKKTKSKKLDGLLERRIKQTELEKQBLEKI 1069
DB 1599 COOHTSHSQSFSLVDI-----KOKIAGNSYVSLEEFYDMISOVIOQ- 1639
QY 1070 KLEGGIKIGKGTSTSSKNLSESPVITKAQEGQSDMRQESPNANNDQ-----PED 1122
DB 1640 -----SNCDELDTAYK-----ELLSEQFPFQNETKACTDALEED 1674
QY 1123 LIQGCC-----QSDSSVLRMSDPSHTTNKLYPKDRVLDVYSIRSPETKCPQKNSIEND 1175
DB 1675 MFESCSGGVYEDLQDAGGVASVYNHSTSQAESRSGVLD-----IPL----- 1717
QY 1176 IEKYSVDLASRQOEPTKSKTKGNDPFIDDDSKLASADDITLICKNK-----KPLI 1225

```
Db 1718 --EEVDGSGGCIK-MRLDTRMCLFCRKSGEGLSGEEARLLYCGHDCWHVHTNCAMWSAEV 1774
QY 1226 QEESDTIVSSSKSAHSSVPKS-----TNRDATPLSRAMDFEGKLGCDSDS-----N 1273
Db 1775 FEED-----GSLQNVHSAVAGRMKCTVCGNRAT-----VGCNVRSCGHHY 1819
QY 1274 STLENSSDTVSIODSS-----EEDMIVONSNEISSEOFRTREODVEVLEPLKCLVS 1325
Db 1820 YPCARSIDCAFLTDKSMYCPAHAKNGNALKANGSPSV-----IYESNFEVSRPVYVEL-- 1872
QY 1326 GESTGNCEDRLPVGTGTEANGKPSQOK-----KLEER-----PVN-KCS 1363
Db 1873 -----DRRKKLLIE-----PARVQFHGSLSEVRQLGAIIVPRFSDSYEAVVPINFLCS 1919
QY 1364 -----DQIKLKNITDKNNENRESE-----KKQRTSTFOINGKDNKPKIYLKGECLKE 1412
Db 1920 RLYSSKEPWKIVEYVTRTTIONSSSILTALDVGRTVTDHTNPSNEKEVQLGMAIARWH 1979
QY 1413 ISESR---VVSNGVE-----PKVNNINKIIPENDIKSLTVKESAIRPFINGDVIMEDFNE 1464
Db 1980 TSLARSEPLENGGTDWSEFPNPSNC--VPDQNTSEEPQOQADLLPPELKDAIFEDL-- 2035
QY 1465 RNSSETKSHLLSSDAEN---YRDSLETLPSTKES-DSTQTTPPSASCPSNSVNOVED 1520
Db 2036 -----PHELDDGISMLDIFLYDKTDLFAISQSKDGTQAMTS-----NOAQN 2078
QY 1521 MEIETSEVKKVTSSPITSEESNLSNDFIDENGLPI-KNKNENVAGESKRRKTVITEVTMT 1579
Db 2079 QNOQAGGANSVS-----ICDETRNSNTSLG-NGWPASNVEDAMLSAARN---SSQVQML 2130
QY 1580 STVA-----TESKTVIKVEKGDKQTVVSTENCASKSTVTTTTVTVKLSTPTSGGSVD 1632
Db 2131 KTLAWPKLDGNSAMATAIKRRLSK-----NLAEGVFTLSSQQRKEMATVAGV- 2181
QY 1633 IISVKEQSKTVTTVTDSLTTCGTGLVTSMT--VSKEYSTRD-----KVKLKFSRP 1683
Db 2182 -----SRQISSETSEVEGATVTSVRSKSTFWSAAKRYFEKSGREBAKRMQMDG- 2235
QY 1684 KTRSGTALPYRKPVTKSTKKSIFVLPNDLKLARKGIR-----EVPYFNYNAPALD 1739
Db 2236 -----VDDSITEFRIISGDGNLSLTAQFSQGVKCDRCQCTRYNYDAFOR-- 2278
QY 1740 IWPYSPRPTGIIWRYLQIVKSLAGVSLMLRLLLWASLRWDDMAAKVPPGGSTRFETS 1799
Db 2279 --HLPSCSPT-----MSSNETESDVS-----GQGTNNATQ 2307
QY 1800 ETEITTTTELKRRDVGPIRGFEYCIRKIICPIGVPEPKETPTQR-----KGLR 1850
Db 2308 ISAESLNEQLKOLLANAGLNY-----LQSATSPQVORLGSLGQFGLQLQ 2354
QY 1851 SSALRPK-----RPETPKOTGPVILTIETWVAEELEW--EIRAFARV----- 1891
Db 2355 QLQLQPSGLNGFFLSQPNPATQANT-----DDLQIYANSLQGLAANLGGGFTL 2403
QY 1892 -----EKEKAQAEQQAQKLEQKQKTVIATSTTSPTSTSTISPAQKVMVAPI 1941
Db 2404 AQPTVTAPAQPOLIAVSTNPQGTQOFIQIPTMQATTTPTATYQTLQATNTDKKIML--- 2460
QY 1942 SGSVTTGKMWLT--TKVGSFATVTFQONKPHOTFAFWK-----QGCNSGVVQVQO- 1993
Db 2461 --PLTAAGKPLKTVATKAQAQAAVKQRLKSGHVKPQLQAKLPHQPHQHQOQQTQVQP 2518
QY 1994 -KVLG-----IIPSSTGT-----SQQFTSF----- 2013
Db 2519 ITVMGNLLQPOLLPQSSTQTAQPIILPQAQPNIIISFVTDGSGQGPLOAYISIPTAGE 2578
QY 2014 ---QPR-TATVTIERTSSGG-----TTSNS--QVITGPOLRPGMTVIR 2052
Db 2579 YKPQOPATPFLTAPCAGATYLOTASGNLVLTTPSNSGLQMLTAQSLQAQPVIG 2638
QY 2053 TPLQOSTL-----GKAIITPVVMVQPCAPQVM-----TQIIRG 2086
Db 2639 TLIQPTIQLGGGADGN-----QPGNQOPLILGGTGGSSGLEFATTPSQVILAT 2689

2087 QY -----STAVSAPNTVSST-PG-----QKSLTSATSTSIQSSASOPPRPQOGVK 2133
Db 2690 QPMYUGLETIVQNTVMSQQOFVSTAMPGLMSQNASFSATTTQVFOASKIEP-----IVD 2743
QY 2134 LTMQALQTOLTOGGHNOGLTV-----VLOGGO--TTGQLOLIPQGVTVLPGPQQLMQAA 2187
Db 2744 LPAGYVVLNNTGDASSAGTFLNAASVLQOQOTDDTTQI----- 2782
QY 2188 MPNGTVORFLTPLATTATTASTTTTSTVSTAAGTGEORQSKLSPQMQVHQDKTLPPAQS 2247
Db 2783 LQANFQ-FOSVFTSSGASTSMDYTSPVMYA-----KIPPVTOIKRTNAQAKAAG 2832
QY 2248 -SSVGPAKAQO-----TAQPSARPOQOTQPSQPAQPEVQTOPEVQTOFTVSSHVPS 2299
Db 2833 ISGVGVPPQPVVNVKVLPTSIVTQSQOVQVKNLKSQVKGKAASGTGTTCGA-PPSI 2891
QY 2300 A-----OPTHAQSSKPOV-----AAQSQPOSNVQGSQSPVRV--Q 2331
Db 2892 ASKPLQKKTNMRIRPIHKLEVKPKVMKPTPKVQNONHSLLOQOQOQPOLQOQIPAVVNVQ 2951
QY 2332 SP-----SOTRIRPST-----PSQLSPGOOSQVOTTTTSQPI-----PIQPHTSLQ 2371
Db 2952 VPKVTISOQRIPACTOQOQLQQAQMIHIPOQOQPOLQOQOVQVQSPMPITLAEAPVVQSQ 3011
QY 2372 IPSQGPQSQPOVOSSTQTLSSSGTFLNQVSVSSPSR----- 2407
Db 3012 FVMEPQALEQOELANRVQHFSTSSSSSSNCSLPTNVVNPQMQAQFSTTSSSTRPTNRV 3071
QY 2408 -POLQIOQPOP-----QVIAVPOLQOQOV--VLSQIQSQVVAQITQAQSGVGPQOIKLQLP 2459
Db 3072 LPMQORGPAPLSNECPVSPPTPKPVEQPIIHMQTSASVSKYAGKSTLPSPV---YE 3128
QY 2460 IQIQSSAVQTHQIQNVTVTVOAASVQOL-----QRVQOOLRDQOQKKK 2502
Db 3129 AELKVSSVLES--IVPDVTMDAILEEQPVQTSIYTEGLYEKNSPGESKTEQLLLQOQORE 3186
QY 2503 Q--QOIEINVTFSKLLIKVEIIOKVVMKHNVAIEH-----LKOKKS-MTPABREEN 2552
Db 3187 QLNQOVLNNGYLLDKHFTQVEPMDTDVYREEDLEEEDDDDFSLKMRTSACNDHEMSDS 3246
QY 2553 QRMIVCNQVMKYILDKIDKEE-----KQAQKKRK 2581
Db 3247 EEPVAKDKISK-ILDNLNTDDCADSIATATTMEVDASAGYQOMVEDVLATTAQASAPTEE 3305
QY 2582 REESVEQKRKQNKATKLSALLFKHKEQLRAELKRALLD-KDLOIYQVQELKRDLIKK 2640
Db 3306 FEGALETAAVEAAATYINEMADAH-----VLDLQQLQNGVELELRRR---KE 3349
QY 2641 EKDLMLQALQATAVAPCPVPTVLPAPPAPPPPPPPG 2679
Db 3350 EORTVSOEEOGSKAIVP--TAAAPPEPQIQEPKKMTG 3386

RESULT 3
PCLO_RAT
ID PCLO_RAT STANDARD; PRT; 5085 AA.
AC Q9JKS6; Q9JLTL;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN PCLO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX MEDLINE=20170257; PubMed=10707984;
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a presynaptic zinc finger protein structurally related to
```

RT bassoon";
 RL Neuron 25:203-214(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
 RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
 RP ALA-4694.
 RX MEDLINE=21181819; PubMed=11285225;
 RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
 RT "An unusual C(2)-domain in the active-zone protein piccolo:
 RL Implications for Ca(2+) regulation of neurotransmitter release.";
 EMBO J. 20:1605-1619(2001).
 CC -1- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -1- SUBUNIT: Interacts with Rabac1/Pral and profilin.
 CC -1- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
 CC junctions.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9UKS6-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9UKS6-2; Sequence=VSP_003930, VSP_003931;
 CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -1- SIMILARITY: Contains 2 C2 domains.
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF138789; AAF07822.2; .
 DR EMBL; AF227534; AAF63196.1; .
 DR HSP; P04410; 1A25.
 DR GO; GO:0045202; C:synaptic junction; IDA.
 DR GO; GO:0005509; F:calcium ion binding activity; IDA.
 DR GO; GO:0005544; F:calcium-dependent phospholipid binding acti. . . ; IDA.
 DR GO; GO:0005522; F:profilin binding activity; ISS.
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO; GO:0016080; P:synaptic vesicle targeting; NAS.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00168; C2; 2.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00239; C2; 2.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS00499; C2_DOMAIN_1; 1.
 DR PROSITE; PS50004; C2_DOMAIN_2; 2.
 DR PROSITE; PS50106; PDZ; 1.
 KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
 KW Repeat; Alternative splicing.
 FT DOMAIN 372 491
 FT 12 X 10 AA TANDEM APPROXIMATE REPEATS OF
 FT P-A-K-P-Q-P-Q-P-X.
 FT C4-TYPE (POTENTIAL).
 FT C4-TYPE (POTENTIAL).
 FT POLY-PRO.
 FT PDZ.
 FT C2 DOMAIN 1.
 FT C2 DOMAIN 2.
 FT TKPTN -> SKRRK (in isoform 2).
 FT /FTID=VSP_003930.
 FT VARSPLIC 4876 4880
 FT Missing (in isoform 2).
 FT /FTID=VSP_003931.
 FT VARSPLIC 4881 5085

FT	MUTAGEN	4668	4668	D->A: COMPLETE LOSS OF CALCIUM-BINDING AND CALCIUM-DEPENDENT PHOSPHOLIPID BINDING ACTIVITY.
FT	MUTAGEN	4674	4674	D->A: COMPLETE LOSS OF CALCIUM-BINDING AND CALCIUM-DEPENDENT PHOSPHOLIPID BINDING ACTIVITY.
FT	MUTAGEN	4688	4688	V->S: SMALL INCREASE IN AFFINITY FOR CALCIUM.
FT	MUTAGEN	4689	4689	VM->SS: 10-FOLD INCREASE IN AFFINITY FOR CALCIUM.
FT	MUTAGEN	4689	4689	M->S: INCREASED AFFINITY FOR CALCIUM.
FT	MUTAGEN	4690	4691	VV->SS: 10-FOLD INCREASE IN AFFINITY FOR CALCIUM.
FT	MUTAGEN	4692	4693	QN->AA: MODERATE INCREASE IN AFFINITY FOR CALCIUM.
FT	MUTAGEN	4694	4694	A->S: NO EFFECT ON CALCIUM-BINDING ACTIVITY.
FT	SEQUENCE	5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;		

Query Match 3.0%; Score 451; DB 1; Length 5085;
 Best Local Similarity 18.1%; Pred. No. 2e-08;
 Matches 587; Conservative 445; Mismatches 1101; Indels 1116; Gaps 139;

QY	276	ECVKPPLEVPDEWQCEVCVAHKVP	-----GVTDCVAEIQKNKPIYRHEPI	322
DB	1117	EAEPK-----VPEK-----ETASIEKTPPMVTTDQKLESEGKSKVSALPEKPKSEEKAI	1168	
QY	323	GYDRSRKRYWFLNRLLIIEEDTENENEKKIWIYSTKVQLAELIDCDKDY-WEAE-LCKI	380	
DB	1169	SADKKERK-----PPAEKPPLEKKPI-----PVDKLLPPEAKPLSSE	1207	
QY	381	LEEMREEIHRHMDITEDLTNKGKSNKSFIAAANEIILESIRAKKGIDIDN-----	430	
DB	1208	GEEKHEILKAHVOIPE-----EPTGKVAAGAGEEQDPSRPEALP	1249	
QY	431	-----VKSPETEKKD-----NETENDSKDAEKNREEDFQSLKSDDDKTPDDDEQKSE	482	
DB	1250	GATPLPLPKAGEKERAVAQPAEGSKDQGGERSK---EKTEKE-EDKSDTSSSQPKSP	1305	
QY	483	EP-TEYGDGKNSVANLGNNTT-----NATSEETSSEGE-----RSPVG	520	
DB	1306	QGLSDGTGYSDGSGSLGEIPSLIPSEKDLLKGLKDSFQSESSPSDLAKLESTVL	1365	
QY	521	CLSETPDSSNMAEK-----KVASELPQD-----VPEE-----	547	
DB	1366	SILEAQASTLVGEKAEEKTOPQKISPEKPODQQTQTASSETLDTITSEEEIKESQEKKVS	1425	
QY	548	PNKTCES-----SNTSATTTISIQPNLENSN-----SSELNSSQS	582	
DB	1426	PKKDEQGFPSRKEHKEKPELVDDLSPRASYSVDSSESENSESVVRRKRRTSIGSSS	1485	
QY	583	ESAKAADDPENGERSHPTPVSIQEIIVGDFTSKSTG-----ELS-----	622	
DB	1486	DEYKQEDSQSGEEDF-----IRKQIEMSADEDSGEDEEFIRSQLKEISGVGSQKR	1541	
QY	623	-ESPGAGKAGSGSTRIITRL-----RNPDSKLSQLKSKQOVA	657	
DB	1542	EEAKGKGKGVAGKHRLTRKSTSFDDAGRRHSHWDEDETTDESPELKFRETQSQE--	1599	
QY	658	AAAHAEANKLFKEGKVLVNVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKYRYVINOYS	717	
DB	1600	-----SEELVVGAGGGLRRFKT---IELNSTIAD-----KYSSESQKK	1635	
QY	718	TNSFALNKQHREDHDKRRLAHKFCLTTPAGEPKWNG-----SVHGSKVLTISLRITITQ	773	
DB	1636	TILY-----FDEEPELEME-SLTDSPEDRSGEGSSSLHASS-FTPGTSPTSVSS	1683	
QY	774	LENNIPSSFHPNASHRANWIKAVQMSKCPREFALALAILCAVAPVVMPLPIWREFLGH	833	
DB	1684	LDSDSDSPSHKKGESQ-----QKARHSHGPLLTIE-----	1718	
QY	834	TRLHRMTSTERE---EKEKVKKEKQEEETMQATWVYTFPVKHQVWKQKGGEYRVT	890	

Db 1719 -----DSSEBELREBELLEKEQKQRELEQOQKQSSKSKKDDKDLRAQRERR--- 1769
Qy 891 GYGWSWISKTHVRFVFKLPNGTNNVYKSLGKTN-NMDENMDESDDKRCR- 943
Db 1770 -----PKTP-PSNLSPIDASTFEELRQAAEMBELHRSSCSYSPSIE 1811
Qy 944 -SPKKIKIEPOS--EKDEWGSDAAGA-----DONEMD-----ISKITEKKDDQKELLD- 991
Db 1812 SDPEGEFISPEIKIEVQVKYKLPAAVSLYSPDQSVNQKQEGVQKALKSABEYEMMQK 1871
Qy 992 -----SDSKPKKEEPMEDDDMKTESHVNCQESSQOVVVNVY-EGFHLRTSYKK 1040
Db 1872 PHKYKAPPAANERDEVFEPLY--GGMLIEDYI--YESLVEDYNGSVGSL- 1921
Qy 1041 KTKSKLDGLLERIKOFTLEKORLEKIKLEGKIGIG-----KTS 1082
Db 1922 -TROEQNGFMQGRG---EOKYRLOEOIYDDPMQKISDLQKEFEYESLHVVPOEDI 1976
Qy 1083 TNSKNKJESPIV-----TKAEGCOSDS-----MRQOSPNNANDQPEDLIOGC 1127
Db 1977 VSSVYIPESHEIVDLGSMVMSTSEKKLLDADSAYELMRQOVQVTDGSSPVQTTIGD 2036
Qy 1128 SQDSSV--LWSDPSHTNKLYPKDRVLD----VSIRSPETKPKONSTENDIEKVS 1181
Db 2037 DMAESTLDFRVQDASLTSSIL--SGASLTDSATLSIPDKITQOFSAELEDEYVT 2094
Qy 1182 DLASRGQBPTKSKTKGNDFIDDKLASADDIGTLCKNKKPLIQEESDTIVSSKSAH 1241
Db 2095 DYTREIQD-----IHAESLIITYSEPSESA 2120
Qy 1242 SSVPKSTNDRDATPLSRAMDFEGKLGCDSESNSTLENSDTSVTSQDSEEDMIYQNSNES 1301
Db 2121 TSVPPSDPPLSTSSISSV-----CTTDSPPVTLTLDSTLVYTEPADVMYTFKDSSE 2172
Qy 1302 ISEQPTRE---ODVEV-LEPLKCELVSGESTGNCEDRLPVKGTANGKKPSQOKKLE 1355
Db 2173 ISSYTFPGSIIDYDEIVSULD-----RTIMPESRTN-EDRVL---SFSGMAPSVVESVG 2224
Qy 1356 ERPVNCSDQIKLNTDKKNENRESEKKQORTSTFOINGKDNKPKIYKLGCEKKEISE 1415
Db 2225 TKPERQADTI---STDLPISE--KOLIKGK-----KETGDGIL--EVLDAYKD 2267
Qy 1416 SRVSGNVEPKVNNINKIPE-----NDIKSLTVKESAIRPF-INGDVIMEDFNENS 1467
Db 2268 KREES---EALTKIS--LPEPGLAQAPSSVTAPOIKEQHVSPHSVSGKI-----SG 2314
Qy 1468 SETKSHLLSSDAEGNRDSLETLPSTKESD---STQTTTPSASCPESNSVQVEDMEIE 1524
Db 2315 QEKPTYRLPSGLPVSTHPS-KSRFFFRSSSLDISAQPPPPPPPPPPSPS----- 2363
Qy 1525 TSEVKKVTSSPITSEESNLNDFIDENGLPINKNENYNGSKRTVITEVTTMTSTVAT 1584
Db 2364 -----TSSPPPTP-----LPPATSPKPTYPKKKLAVAAVTTSTIVTT 2403
Qy 1595 --ESKTVIKVEGDKQTVVSSTENCA-----KSTVTTTTTIVT-----KLSPTSGGS 1630
Db 2404 HYDALTMYEAAARSNSGLPATKCAIAPPVPPPKPSQIPTGLVTHRPEAIKPP----- 2458
Qy 1631 VDIISVBOQSQTWVTTT--VTDLSLTGTGLVMTSVSKYSTKDRKVKLMKFSRPPKTRS 1688
Db 2459 ---IAPKPAVQIPVYTKQPTDTCPKPTGLSLTS-TMSLNLTSDYNNVPSTPLSPHS 2514
Qy 1689 GTALPSYKFKVTSTKKSIFVL-----PNDDLKKLARKGGIREVPYFNYNKAPALDI 1740
Db 2515 NKSSPRY----SKSLMDTVVITLPSPEGTPDTS-----SAAQAITS 2552
Qy 1741 WPIPSRPTFTGITHRYRLQTVKSLAGVSLMLRLLWASLRWDDMAKVPPPGSGSRTERISE 1800
Db 2553 WPLGSPPKDL-----VSLTFVSV-----VPPMTSTEIPSAQ 2585
Qy 1801 TEITTTETIKRRDVGp--YGRIFEYCIKRIICPIGVPE-TPKEPTPORKGLRSS--ALR 1855
Db 2586 PTLTYSGLGTFSTVPATVASLFQ-----TVPSLTQFLPAEASKPEVSAVSAVPSVA 2639
Qy 1856 PKRPETP-----KQTG--PVI-----IETWVAEEELWELWIRAPAE 1899
Db 2640 PRSVSIPPEPLALDRHQYKENGKPLIGDAIDLRTIPKSEKVKTEKMDL-----SAS 2694
Qy 1890 RVEKEKAQAEQOAKKLEQOKPVIATITSPSSSTSTISPAQKVMVAPISGVSVTGT 1949
Db 2695 AMDYKROTANEVYRQOISAVQPSIIINLSAASSLGTPTVMDSKTVAVVYCTDTTIYTTGT 2754
Qy 1950 --KWLVTTKVGSPTAVTFQON-----KNFHOTFATVWKQGSNSGVVQOQKVLGIPS 2001
Db 2755 ESQVGIHATVSPLOLITTSKHTLPYKPPSSQAEPPTIRDEAPIN-----L 2799
Qy 2002 STGTSQOFTFSQPRATVATVIRPNTSGSGTTSNSQVITGPQI-----RGMVTVIRPLQ 2056
Db 2800 SLGPSAQAVTLAVTKPVTPPVGVNTGTSTLSQGVADGEVDLSTSKSHRTVVTMDSES 2859
Qy 2057 QSTLIGKAIL---RTPVMVQCAPOOVMTOIIRGPOVSTAVSAPNTVSTP----- 2103
Db 2860 TSNVVTKIIIEDEKPVDLTAGR-RAVCCDMVYITLPPFRSCTAQAPATILPDRFGYRDDH 2918
Qy 2104 -----GOKSLTSATSTSNIOSSASQPPRPOQGOVKLTMAQLTQTOCHGNGOGLT- 2153
Db 2919 QYDRSGPYGYRGI-----GGMKPSMSDTNLPAGHFFYKSKNAFDYSGGTGAADVLT 2972
Qy 2154 -----VVIQGGQTTGOL-----QLI-----PQGVV--VLPQGGQOLMQAAM---PNG 2191
Db 2973 GRVSTGEVMDYSSKTTGYPETROVISGVGISTPQYSTARLTTPPPQYGVGSVLRSSNG 3032
Qy 2192 TVQREFLTATTATTATTTTITVSTTAAG--TGEQROSKLSQPMQVHQDKTLPPAQSSV 2250
Db 3033 VVYSSVATPIPTFAITTPQGSIFSTVRLDLSGIPPTDAMTSLTSLHQQSQPMPSRYIT 3092
Qy 2251 GPKAKAQPTAQSPARPOQOTPOQSPAQEVQOTQPTQTVSSHVSEAOPTHAQSKP 2310
Db 3093 GASET-----DIAVTGIDINASLOT-----ITMETLPAETMDSVP 3127
Qy 2311 QVAAQSPQSNVQCGSPVRVQSPQTRIPSTPSQLSQGQSQVQVTTTSQPIQPHSL 2370
Db 3128 TLTATSEVSEFVSGEESTLLIVPDEDK-----QOQQLDL-----ERELL 3166
Qy 2371 QIPSGOPQSPQVQVQSTQTLSSGOTLNOVSVSS-----PSRQLOIQQPOQVIAVPQL 2425
Db 3167 ELEKI KOORFAEELEWERQELQRFREQEKIMVQKKLELOSKQHLLYQOEEERQAQFMM 3226
Qy 2426 QQQVQVLSQIQSVQVQIQAOQSGVQOIKIQLPIQIQ-----SSAVOTHOIQNVVT 2478
Db 3227 ROETLAQOOLQLEQIQLOQOQ---LHQOLEBQKRLQIQVYNDPDSGTSSPOTTTEQAILE 3283
Qy 2479 VQAASVQEQLOQVQOQLRDOQKQKQOQIEINVNTPSKLLIKVELIQK----- 2526
Db 3284 QGYAATE-----GSQFWATEDATTASTVVAIEIPQSGQWTVVQSDGVQ 3328
Qy 2527 -----VVMKHNVAIEHLKQKSMTPAEREENORMIVCNQV-----MKYIIL- 2567
Db 3329 YIAPPGLISTVSEIPLTDVVVKEEKQPKKRSGAKVRQYDEMGESVADDERNLKKIVDS 3388
Qy 2568 --KIDKEE-----KOAKKRKEESVE-----OKRSKQNA----- 2595
Db 3389 GVQTDDEADRSYASRRRTKKSVDTSVQTDDEQDDEWDMPSRSRRKARTGKYGDSTAE 3448
Qy 2596 -----TKLSALL-----FKHKEQLRAEILKRLALLDKDLQ-- 2625
Db 3449 GDKTKPLSKSVSAVQTVAEISVQTEPVGTITRTISARVDAKVEIKHISAPEKTYKG 3508
Qy 2626 -IEVQEELKRLD-----KIKKEKLMOL-----AOATAVAACP----- 2658
Db 3509 SLGCQETDSTDSQPPYLGATSPPKDKKRPPLGIGYSSSHLRADPTVOLAPSPKPKV 3568
Qy 2659 ---PVTPVLP-----APPAPPPS----- 2673
Db 3569 LYSPTISPLSPGNALEPAFVPEVPEKPLPDDISPOKVLHPDMAKVPVPPASPKTAKMMQMSDP 3628

QY 2674 -----pppppgvqhtgllst-pt 2690
 Db 3629 KPIsPTADESSRAFFOYSEGTTKGSQTMASQTOKKVKRTLPNPPPEEYSTGTQSTYST 3688
 QY 2691 LPVASOKR-----KRE---EEDSSSKSKKMIKSTTSKTKKDKL-- 2729
 Db 3689 MGTASRRRCRTWMARAKILQIDRLDIVERESAKLRKKQAEDEEKEI--DAKLRY 3746
 QY 2730 -----YCIKTPYDESKFYI-GCDRCQNYHGRVCVILQSEAEI----- 2767
 Db 3747 LEMGINRRKEALLKEREKREAVLQGVAEEDRY-----MSDSEVSSTRSRVESQH 3797
 QY 2768 -----IDEYVPCQCTEDAMTVLTLTEKDYEGRLRVLRSLOAHKMAWPFLE 2815
 Db 3798 GVERPRTAPQTEFSQIPPTQTEAQLVPTSPYTOQYS-----S 3838
 QY 2816 PVPNDAPDYGVYKPEMDLATWEORVR--YYEKLT-----EFVADMTKIFDNCRY 2867
 Db 3839 PALPTQAPTY-----TOOSHFOQQTLLHQVSPYQPTQPTQAVATMFTQQAQP 3888
 QY 2868 NPDSPPFYQ 2876
 Db 3889 TPTPQPSYQ 3897

RESULT 4
 BA2A_HUMAN
 ID BA2A_HUMAN STANDARD; PRT; 1878 AA.
 AC Q9UIF9; O05336; O15030; O96H26;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Bromodomain adjacent to zinc finger domain 2A (Transcription
 DE termination factor-I interacting protein 5) (TTF-I interacting protein
 DE 5) (Tip5) (bWALp3).
 GN BAZ2A OR TIP5 OR KIAA0314.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-testis;
 RX MEDLINE=20130112; PubMed=10662543;
 RA Jones M.H., Hamana N., Nezu J.-I., Shimane M.;
 RT "A novel family of bromodomain genes.";
 RL Genomics 63:40-45(2000).
 RN [2]
 RP SEQUENCE OF 332-738 FROM N.A.
 RC TISSUE-Lung;
 RA Jansa P., Grummt I.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 639-1878 FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding regions of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 RN [4]
 RP SEQUENCE OF 1035-1878 FROM N.A.
 RC TISSUE-Lymph;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: May play a role in transcriptional regulation
 CC interacting with ISWI. May serve a specific role in maintaining or
 CC altering the chromatin structure of the rDNA locus (By
 CC similarity).
 CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NORC
 CC (nucleolar remodeling complex).
 CC -!- SUBCELLULAR LOCATION: Nuclear. Co-localizes with the basal RNA
 CC polymerase I transcription factor UBF in the nucleolus.
 CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in most tissues
 CC analyzed, including heart, brain, placenta, lung, skeletal muscle,
 CC kidney and pancreas.
 CC -!- SIMILARITY: BELONGS TO THE NAL FAMILY.
 CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
 CC -!- SIMILARITY: Contains 1 methyl-binding (MBD) domain.
 CC -!- SIMILARITY: Contains 1 bromodomain.
 CC -!- SIMILARITY: Contains 1 DDT domain.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB032254; BAA89211.1; -;
 DR EMBL; AF000422; AAB60864.1; -;
 DR EMBL; AB002312; BAA20773.1; -;
 DR EMBL; BC008965; AAH08965.1; -;
 DR HSSP; Q92831; I891.
 DR Genew; HGNC:962; BAZ2A.
 DR MIM; 605682; -;
 DR GO; GO:0005731; C:nucleolus organizer complex; NAS.
 DR GO; GO:0003677; F:DNA binding activity; NAS.
 DR GO; GO:0030528; F:transcription regulator activity; NAS.
 DR GO; GO:0006338; P:chromatin modeling; NAS.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 DR InterPro; IPR000637; AT_hook.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR004022; DDT_dom.
 DR InterPro; IPR001739; Methyl-CpG_bind.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF02178; AT_hook; 4.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF02791; DDT; 1.
 DR Pfam; PF01429; MBD; 1.
 DR Pfam; PF00628; PHD; 1.
 DR PRINTS; SM00503; BROMODOMAIN.
 DR SMART; SM00384; AT_hook; 4.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00571; DDT; 1.
 DR SMART; SM00391; MBD; 1.
 DR SMART; SM00249; PHD; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
 DR PROSITE; PS50016; ZF_PHD_2; 1.
 KW Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;
 KW Nuclear protein; Repeat; DNA-binding.

FT	DOMAIN	525	577	MBD.	
FT	DNA_BIND	622	634	A.T HOOK 1.	
FT	DNA_BIND	643	655	A.T HOOK 2.	
FT	DOMAIN	821	883	DDT.	
FT	DNA_BIND	1159	1171	A.T HOOK 3.	
FT	DNA_BIND	1377	1389	A.T HOOK 4.	
FT	ZN_FING	1649	1699	PHD-TYPE.	
FT	DOMAIN	1783	1853	BROMODOMAIN.	
FT	DOMAIN	633	772	LYS-RICH.	
FT	DOMAIN	666	765	COILED COIL (POTENTIAL).	
FT	DOMAIN	1185	1250	GLU-RICH.	
FT	DOMAIN	1263	1384	PRO-RICH.	
FT	DOMAIN	1732	1735	POLY-ARG.	
FT	CONFLICT	574	574	L -> V (IN REF. 2).	
FT	CONFLICT	700	700	L -> Q (IN REF. 2 AND 3).	
FT	CONFLICT	720	720	H -> Q (IN REF. 2 AND 3).	
FT	CONFLICT	727	738	SKAEKEGKTKO -> KIKKKKKKKK (IN REF. 2).	
FT	CONFLICT	785	785	K -> R (IN REF. 3).	
FT	CONFLICT	951	951	P -> L (IN REF. 3).	
FT	CONFLICT	1005	1006	GR -> EG (IN REF. 3).	
FT	CONFLICT	1035	1037	IAA -> GTR (IN REF. 4).	
FT	CONFLICT	1163	1163	G -> S (IN REF. 1).	
FT	CONFLICT	1166	1166	R -> L (IN REF. 1).	
FT	CONFLICT	1172	1172	S -> F (IN REF. 1).	
FT	CONFLICT	1178	1178	L -> F (IN REF. 1).	
FT	CONFLICT	1202	1202	A -> V (IN REF. 1).	
FT	CONFLICT	1282	1282	P -> L (IN REF. 1).	
FT	CONFLICT	1295	1295	L -> F (IN REF. 1).	
FT	CONFLICT	1313	1313	P -> L (IN REF. 1).	
FT	CONFLICT	1407	1410	MISSING (IN REF. 4).	
FT	CONFLICT	1416	1416	R -> P (IN REF. 1).	
FT	CONFLICT	1541	1541	R -> P (IN REF. 1).	
FT	CONFLICT	1571	1571	E -> K (IN REF. 1).	
FT	CONFLICT	1616	1616	V -> I (IN REF. 1).	
FT	CONFLICT	1622	1622	E -> Q (IN REF. 1).	
FT	CONFLICT	1629	1629	Q -> H (IN REF. 1).	
FT	CONFLICT	1636	1636	Q -> H (IN REF. 1).	
FT	CONFLICT	1739	1739	R -> K (IN REF. 1).	
FT	CONFLICT	1754	1754	G -> R (IN REF. 1).	
SQ	SEQUENCE	1878 AA;	208480 MW;	400970CA68234317 CRC64;	

Query Match 3.0%; Score 447.5; DB 1; Length 1878;

Best Local Similarity 18.6%; Pred. No. 8.9e-09;

Matches 398; Conservative 297; Mismatches 763; Indels 677; Gaps 89;

QY	1082	STNSKNLSVPITKAKGCGQSDMRQSPNANNQPEDLIQGCQSDSVLRMSDPS	1141
DB	103	SANPGSNLKDPPLLSQFSGG-----QYPLNGILGSRPSS-----PS	140
QY	1142	HTTN-----KLY-----PKDRVLDDVSIRSPTKCPKQNSIENDIEEKVSDLASRQOE	1189
DB	141	HNTRLRAGSQFWANGTHSPMGLNFDQELYD---SFPDQ---FEVCS-----GIH	187
QY	1190	PTKSKYGNDFDIDSKLASADDIGT-LICKNNKPLIQESDTIVS-----SKSALH	1241
DB	188	PDEAAEKEMTSVVAEN-----GTGLVCSLE---LEEOPELKMCYNGSVSPVESLH	236
QY	1242	SSVPKSTNDRATPLSRAMDFEGKGLCDSESNTLENSDTSVIOSSSEDMIVQNSNES	1301
DB	237	QEVSVLVPD-----PTVSCID-----DPSHLPOLEDTPILS-----	268
QY	1302	ISEQFRTREQDVEVLPLKCELVSGESTG-----NCEDRLPVKTEANGKKPSQOKKL	1354
DB	269	-----EDSLEPFNSLAPEVPVSGGLYIGDIDPELMAEDKLP-----L	304
QY	1355	BERPVNKCSDQIKLNTDKNNENRESEKKGTSTFTFOINGKDNKPKIYLKGECLKEIS	1414
DB	305	EDSPVTSALDCPSLNNAI-----AFSLUADDSQ-----	332
QY	1415	ESRVVSGNVPKVNINKIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSSETKSHL	1474
DB	333	-----TSTSFASPTSPPVILGESVLQD-----NSFD-----L	359

QY	1475	LSSDAEGNYRDSLETLPTSKESDSTQTTTSPASCSPESNSVNOVEDMEIETSEVKVKTSS	1534
DB	360	NGSDAE---QEEME---TQSSDFPPLTQPA---PDQSTIQLHP---ATSPAVSPTTS	407
QY	1535	PITSEESNLSNDFIDENGLPINKNNVNGESRKVIVTEVTMTSTVATESKTVLKVEK	1594
DB	408	PAVSLVSPAASPEIS-----PEVCPAASTVV---SPAVSFVVS	443
QY	1595	GDKQTVVSSSTENCAKSTVTTTTVTVKLSTPSTGGSDIISVKEQSKTVVTTTTVTDLSLT	1654
DB	444	PASSAVLPAVSLEVPILTASVTSKASDPVSPAAFTASPANKDVSFLETTADVEEITG	503
QY	1655	TGGLTVTSMTV-SKEYSTRDKVKL---MKFSRPKTKRSGT-----AL	1692
DB	504	EGLTASGSDVMRRRIATPEEVRPLQHGWRREVRRIKKGSHRWQGETWYYPGCGKRMQF	563
QY	1693	PSYRKFKVTKSKSIFVLPNDLKLARKGGIREVPYFNNAKPALDINDYPPSPRTFGI	1752
DB	564	PEVIKYLNRNLVHSV-----RRE--HFSF-----SPRMPVGD	593
QY	1753	TWYRLQTVKSLAGVSLMLRLWASLRWDDMAAKVP---PGGSGSTRTSET-----	1801
DB	594	FFEEF-DYPEGLQ-----WVOLSABEIPSRIOAITGKGRPRNTEKAKTKEVPKVK	643
QY	1802	-----EITTEIIRRDVGPYIGIRFEYCIRKILICIGVPETPKETPTQKGLRSSAL	1854
DB	644	RGRGRPPKVKITELLNKTDRP-----LKKLEAOETLNEEDKAKIAKSKMKRMQKVQ	695
QY	1855	R-----PKRPETPKQGPVLIETWAEELWEIRAFARVEKKEKAQAVEQ	1901
DB	696	RGECLTIOGQARNKRQETKSLKHKAKEKKEKTKQEKLEKVKREKKEKVM	755
QY	1902	QAKKRL-----BOQKPTVIATSTTPTSSTTST-----	1929
DB	756	KEKEEVTKAKPACKADKTLATQRRLEERQKQOOLEEMKKPTDMLCTDHPDFSRVP	815
QY	1930	-----ISPAQKVMVAPI---SGSVTTGTGM-----	1951
DB	816	GLTLPFGAFSDCLTIVEFLHSFGVKVGLFDPAKDVPKSLGVLOEGLICOGDSLGEVQDLIVR	875
QY	1952	VLTTKVGSPATVTFQONKNPHQTATVWVKQGSN-SGVVQVQVKVLGIIPSSSTQSO-QT	2009
DB	876	LLKAALHDPGPFYCSOSLILGEKVSIEPLTRDNVSEILRCFLMAYGVPEALCDRLRTOP	935
QY	2010	FTSFQP-RTAVTIRPNTSGSGGT-----TSNSQVITGPDIRGCMVIR	2052
DB	936	FOAOPPOQAAVLAFFPVHELNGSTLIINEIDKTLSSSYKKNKWIVEGLRLLR-----LK	990
QY	2053	TPLOQSTLIGKALIIRTPVMVOPGAPQOVM-----TOIIRGQPVSTAVSAPNTVSTPGOKS	2107
DB	991	TVLAKRT-GRS-----EVMGRPEECGLRRRRSRIMEETSGMEEEEESIAAVPGR-1042	
QY	2108	LTSATSTSNIOSSAQPPRQOQGVKLTMAQLTQLTQGHGNGQLTVVYIOGOGTQOLQ	2167
DB	1043	---GRDGEVDATASSIEPELERIEKLSKROLFFRKLLHSSQMLRAVSLQGDQRYRYW	1099
QY	2168	LIP--QGVTVLPVPGPQQQLMQAMPNGT-----VQRFLETTPLATTATTTTTSVSTA	2219
DB	1100	VLPLYLAGIFVEGTENLVPEEVIKETSLSKVAHASLNPAFLFSMKMLAGSNNTASSPA	1159
QY	2220	ACTGEOROSKLSPOQMVHQDKTLPPAQSSSVGPAKAPQTAQPSARPOQTPQSPQAQPE	2279
DB	1160	RARGPRKIK-----PGSMQPRHLKSPVRGODSEQPQAOLQPE	1197
QY	2280	VQ-----TOPEVQOTVTVSSH---VPSEAQPHQAQSKPQVA-----AQSQPSNVQG	2324
DB	1198	AQLHAPAOPOPOLQLOLQSHKGFLEQGSPLSLGOSQHDLSQSFAFLWSLQTSQSHSLLS	1257
QY	2325	QSPVAVQSPQTRIRPSTPSQLSPGOQOVQTTTSQIP--IQPHTSLQIIPISOGOPQSQP	2382
DB	1258	SS-VLTPOSSPGKLDPA-PSQ--PPEEPDPAESSPDQALWFNISAQMFCNAAPTPPP	1313

```

QY 2383 QVQS-----SQTLSGQTLNOVSVSSPSRPLQI-----QQPQ-PQ 2418
Db 1314 AVSEDTPTSPQQLASSKPMNRSAANPCSP-VQFSTPLAGLAPKRAGDPGEMPSPT 1372
QY 2419 VIAVPOLQQQ--VOVLSQIQSOVVAQIAQOQSGVPQQ-----IKLQLP 2459
Db 1373 GLGPKRGRPPSKFFKQMEGRVLTQTAQP--VPPMCSGWVWIRPEMDMLKALHP 1430
QY 2460 IQIOSSA---VQTHQ--IQNVTVQAA--VOEQLQRVQQ-LRDOQQKKKQOQIENV 2510
Db 1431 RGIREKALHKLHKLHKLHKLHKLHKLHKLHKLHKLHKLHKLHKLHKLHKLHKLH 1490
QY 2511 NTPSKLLIKVEIIQOVVHKHNAVIEHLKOKKMTPAERENQBMIVCNQVMYILDKID 2570
Db 1491 -----LOWVELEQRVIM-----SDLQIRGMTCPSPDSTREDLAYCEHLS-----D 1531
QY 2571 KEKQAAKKRRESVQKRSKONATKLSALLFKHKEQLRAEILKRALLDKDLQIE-VQ 2629
Db 1532 SQEDITWRGRGREGLAPOKLT-TNPLDLAVN-----RLAALQNVVERYL 1576
QY 2630 EELKRLIKKEKDLMLQALQATAVAPCPVTPVLPAPPAPPSPPPPPGVQHTGLLSTP 2689
Db 1577 EPLMPHTEVLEKALLS-----TPNGAPEGTTTETSYEITP 1612
QY 2690 TLPVASKRREBEKSSSKKKKMTTSKES-----TKDKLYCICKTPYDESKFYI 2744
Db 1613 RIRWRQTLR-----CRSAAQVCLGLQGLERSTAWERSVANKVTLVCRKGDNDFFLL 1665
QY 2745 GCDRCQWYHGRCVGILQSAELIDE--VVCPOC--QSTEDAMT----- 2784
Db 1666 LCDGCDRCGHYCH--HRPKMEAVPEGDFWCTVCLAQVGEGETQKPGFPKGRKRSY 1722
QY 2785 -----VL-----TPLTEKDY--EGL---KR-----VLRSL 2804
Db 1723 SLNFSEGDGRRRRVLLRGRESAAGPRYSEGLSPSKRRRLSMRHHSDLTFCIILMEM 1782
QY 2805 QAHKMAWPFLEVPDNDAPDYGVVKEPMDLATMEERVQRVRYEKLTFEVADMTKIPDNC 2864
Db 1783 ESHDAWPFLEVPDNDAPDYGVVKEPMDLATMEERVQRVRYEKLTFEVADMTKIPDNC 2864
QY 2865 RYINPDSPPYQCAEVLESFVOKLKGFKASRSHN 2899
Db 1843 QTNEDDSEVKGAGHIMRRFFESRWEFFYQKQKAN 1877

RESULT 5
PCLO_CHICK
ID PCLO_CHICK STANDARD; PRT; 5120 AA.
AC Q9PU36; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Piccolo protein (Aczonin) (Fragment).
GN PCLO OR ACZ.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99439764; PubMed=10508862;
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA Kilimann M.W.
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin."
RL J. Cell Biol. 147:151-162(1999).
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).

```

```

CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions (By similarity).
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y19187; CAB60725.1; -.
CC HSP; P04410; I25.
CC GO: GO:0045202; C:synaptic junction; ISS.
CC GO: GO:0005509; F:calcium ion binding activity; ISS.
CC GO: GO:0005544; F:calcium-dependent phospholipid binding acti. .; ISS.
CC GO: GO:0005522; F:profilin binding activity; ISS.
CC GO: GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
CC GO: GO:0016080; P:synaptic vesicle targeting; ISS.
CC InterPro: IPR000008; C2.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001565; Synaptotagmin.
CC Pfam: PF00168; C2; 2.
CC Pfam: PF00595; PDZ; 1.
CC PRINTS: PR00399; SYNAPTOTAGM.
CC SMART: SM00239; C2; 2.
CC SMART: SM00228; PDZ; 1.
CC PROSITE: PS00499; C2_DOMAIN_1; 1.
CC PROSITE: PS00004; C2_DOMAIN_2; 2.
CC PROSITE: PS0106; PDZ; 1.
CC Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
CC Repeat.
KW NON_TER 1 1
KW DOMAIN 258 357
FT FT 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT P-A-K-P-Q-O-P-X.
FT ZN_FING 368 392 C4-TYPE (POTENTIAL).
FT ZN_FING 836 859 C4-TYPE (POTENTIAL).
FT DOMAIN 2324 2343 POLY-PRO.
FT DOMAIN 4414 4493 PDZ.
FT DOMAIN 4627 4726 C2 DOMAIN 1.
FT DOMAIN 5003 5094 C2 DOMAIN 2.
FT SEQUENCE 5120 AA; 560751 MW; A658D9891B65B412 CRC64;

Query Match 2.9%; Score 441.5; DB 1; Length 5120;
Best Local Similarity 17.8%; Pred. No. 4.4e-08;
Matches 503; Conservative 401; Mismatches 979; Indels 939; Gaps 114;

QY 401 KARGSNKSFLLAANEET---LESIRAKKDDIDNVKSPETEKDKNKTENDSKDAENRE 456
Db 768 KAAPTTKKETKPLASEKLGPMASDSTLTTKGS-DLEKKPSLAKDSKHQTAEPALSEQ 826
QY 457 EFEDQ-----SLEKSDSDKTPDDDP----- 477
Db 827 EKASQPKVSCPLCKTGLNIGSKDPPNPNTCKKKVVCNLCGNPMPHIVEVQEWCLNC 886
QY 478 QGKSEPTVEYDGNKSVSANLGDNTTNTATSETPSEGRSPVGLSTPTDSSNAEKVA 537
Db 887 QTQAMSGQLGDMGKVPPLPKLGFSPQ--VSKPPATPQ--KQPVAVSHSPQKSTPTTPAA 943
QY 538 SE-----LPQDYPEEPNPKTESNTSATTSTTOPNLENSNSSE-----LNSSQSSSAK 587
Db 944 TKPEEPSVPKEVPLKQQGKLEK---TSLADKIQQGQTKEDAKSKGKLFKTPSADKIQR 1000
QY 588 ADDEPENGERE-----SHTPVV-----IOEIVGDFTESEKSTGELSPGAGKSGASTRI 637
Db 1001 VSKEDSRLOOTKLTTPPSDKTLHGVRKEDI-----KFOEAKLAKIPSADK-----I 1048
QY 638 ITRLRNPDSKLSQLKSQVAAAAAHEANKLFKEGKVLVNSQGEISRLSTKKEVIMKNI 697

```


Qy 2375 QGQBPQ-----SQPVQSSSTLTSSGQFLNVOVSSPSRPOLIQO-----POPQ 2418
Dy 2975 YSOARMVSSLSPPGAGSVLRSSNGVSVVATPSTFAITTPGSGFSTVTRDLTLQ 3034
Qy 2419 VI-AVPOLO--QQVQVLSQTSQVVAQIQAOQSGVPOQIKQLPIQIQSSAVQTHQIQN 2475
Dy 3035 TIDVPSLSLQOQNPRLPSYSFLTTWAAEKDASTTLDIETGLPPLTLESIAIEP-----TN 3091
Qy 2476 VVTVQAASVOEQLRVQOQLRDQ-----QOKKKQOQIEINNVTPSKLLIKVIEIQKQVWM 2529
Dy 3092 LIPATASEVYTDV-----IEDEVALIITAPEGKQOQLDL-----ERELLELEKIKQO--- 3138
Qy 2530 KHNATIVHLKOKKSMTPAEERENQRMIVCNQVMKYIILDKIDKEKQAAKKRKRREESVEQK 2589
Dy 3139 ---RFABELEW-----ERQEQRF-----REQEKEMVQK 3164
Qy 2590 RSKQNAFKLSALLFKHKEQLRAE--ILKKRALLDKDLQIE-----VOELKRLDKLKKKEDL 2644
Dy 3165 KLEELQSMKHLLFQEEERQOQYMMRQETLAQOQLQLEQFQOQLQOOLLEEQKIRQI 3224
Qy 2645 MQ 2646
Dy 3225 YQ 3226
RESULT 6
ID K167_HUMAN STANDARD; PRT; 3256 AA.
AC P46013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Antigen Ki-67.
GN MKI67.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043435; PubMed=8227122;
RA Schluster C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,
RA Flad H.-D., Gerdes J.;
RT "The cell proliferation-associated antigen of antibody Ki-67: a very
RT large, ubiquitous nuclear protein with numerous repeated elements,
RT representing a new kind of cell cycle-maintaining proteins.";
RL J. Cell Biol. 123:513-522(1993).
RN [2]
RP SEQUENCE OF 1-31 FROM N.A.
RA Gerdes J.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL
CC PROLIFERATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1
CC PHASE IN THE PERINUCLEAR REGION, IN THE LATER PHASES IT IS ALSO
CC DETECTED THROUGHOUT THE NUCLEAR INTERIOR, BEING PREDOMINANTLY
CC LOCALIZED IN THE NUCLEAR MATRIX. IN MITOSIS, IT IS PRESENT ON ALL
CC CHROMOSOMES.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P46013-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P46013-2; Sequence=VSP_004298;
CC -!- DEVELOPMENTAL STAGE: EXPRESSION OF THIS ANTIGEN OCCURS
CC PREFERENTIALLY DURING LATE G1, S, G2, AND M PHASES OF THE CELL
CC CYCLE, WHILE IN CELLS IN G0 PHASE THE ANTIGEN CANNOT BE DETECTED.
CC -!- SIMILARITY: Contains 1 FHA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; X65550; CAA46519.1; -;
DR EMBL; X65551; CAA46520.1; -;
DR EMBL; X94762; CAA64388.1; -;
DR PIR; A48666; A48666.
DR Genew; HGNC:7107; MKI67.
DR GK; P46013; -;
DR MIM; 176741; -;
DR GO; GO:000074; P:regulation of cell cycle; TAS.
DR InterPro; IPR00253; FHA.
DR Pfam; PF00498; FHA; 1.
DR SMART; SM00240; FHA; 1.
DR PROSITE; PS00006; FHA_DOMAIN; 1.
KW Cell cycle; Antigen; Nuclear protein; ATP-binding; Repeat;
KW Alternative splicing; Polymorphism.
FT DOMAIN 27 76
FT DOMAIN 1000 2928 16 X 122 AA APPROXIMATE REPEATS.
FT REPEAT 1000 1112 1.
FT REPEAT 1122 1234 2.
FT REPEAT 1244 1356 3.
FT REPEAT 1366 1477 4.
FT REPEAT 1487 1598 5.
FT REPEAT 1608 1720 6.
FT REPEAT 1730 1842 7.
FT REPEAT 1851 1964 8.
FT REPEAT 1974 2086 9.
FT REPEAT 2096 2204 10.
FT REPEAT 2214 2326 11.
FT REPEAT 2335 2447 12.
FT REPEAT 2457 2569 13.
FT REPEAT 2579 2689 14.
FT REPEAT 2699 2808 15.
FT REPEAT 2818 2928 16.
FT NP_BIND 3034 3041 ATP (POTENTIAL).
FT VARSPPLIC 136 495 Missing (in isoform Short).
FT VARIANT 3150 3150 /FTid=VSP_004298.
FT VARIANT 3217 3217 T->S (IN dbSNP:111106).
FT VARIANT 3217 3217 /FTid=VAR_014858.
FT VARIANT 3217 3217 K->E (IN dbSNP:8473).
FT VARIANT 3217 3217 /FTid=VAR_014859.
SQ SEQUENCE 3256 AA; 358741 MW; 578F8C51BED42517 CRC64;
Query Match 2.9%; Score 439.5; DB 1; Length 3256;
Best Local Similarity 18.3%; Pred. No. 3.1e-08;
Matches 579; Conservative 432; Mismatches 1103; Indels 1049; Gaps 148;
Qy 300 VPGVTDCAVEIQKPKYIRHEPIGYDRSRKRYWFLNRRLIIEEDTENENEKKIYW----- 354
Dy 221 VP-TTQCLDNSKKNE-----SPFWKLYESVKKELDVKSQENVLQYCRKSG 265
Qy 355 ----YSTKVQIAELIDCKDYWEAEICKILEEMREEIHRHMDITDITNKARQSNKSL 410
Dy 266 LQTDYATEKESADGLQ-----GETQLL-----VSRKSRPKSGSGGHAVA 304
Qy 411 AAAN-EETLESIRAKKGDIDNVKSPET-----EKQKNET-----ENDSKDAEKNR 455
Dy 305 EPASPEQLDQNKGRDVSQVTPSKAVGASFFLYEPARKMKTVPQVYQQQNSQPKHKNK 364
Qy 456 EEF-----EQDSLEKD-----SDDKTPDDPQGGKSEPTVEYDKNQSVS--ANLGDNTN 504
Dy 365 DLYTTGRRSVNLGKSECFKAGDKTLTPRKLTNRTPAKVEDAADSAATKPKENLSSKTRG 424
Qy 505 A--TSEETSPSE---GRSPVGCLET-----PDSSNMAEK-----KVASELPQDVPE 546
Dy 425 SIPTDVEVLPETETIHNPEFLTLWTQVERIKQDSLKPEKLGTTAGQMGSLPGLSSV 484
Qy 547 EPNKTCSSNWS-----ATTTSIQPNL--ENSNSSSELSSQSESAAK----- 588
Dy 485 DINNFGDSINESEGIPLKRRRVSGFGLHRLPELFDENLPPTPLKRGAPTAKRKSLVMHTP 544


```

Db      2384 GKAMDTKPAVSDEKNINFTVFVQKLDLGNLPGSKRQPPPKKAEALDLVGPKEK 2443
QY      2383 -QVSSQTSLSSGQTLNQVSVSPSRPQLQIQPQPQVIAPV---QLQQQVQVLSQI-- 2435
Db      2444 FQTPGHTEESMTDDKITEVSKSPQSPESFKTSRSSKQRLAIPLVKVDKKEEPLAVSKLTR 2503
QY      2436 OSQVVAIQAGQSGVPOQIKL--OLPQIOOSSAVQTHQIONVTVVQAASVQSOLQRVQQ 2493
Db      2504 TSGETTQTHTEPTGDSKSIKAFKESPKQILDPA--SVTGSRRQLRTRKKEKARA 2555
QY      2494 LRDOOQKKK-----OOQIEINVT-----PSKLLIKVELIQ-----KOVVMKH 2531
Db      2556 LEDLVDFELFSAPGHTESMTDKNKIKPCKSPPELTATSTKRCPTTRPKVEKEE 2615
QY      2532 NAVIEHLKOKKSMT-----PAPREENQRMIVQNMVKYILDKIDKEEQAARKRREES 2585
Db      2616 LSAVERLTQTSQSTHTHKEPASGDEGKVL-----KQRAKKKPNPVE 2658
QY      2586 VEQKRSKQNA-----TKLSALLFKHKEQLRA-----EILKKRA 2618
Db      2659 EEPSSRRPRAPKEKAQPLEDLAQTELSETSGHTQESLTAGKATKIPCESPPLEVVDTTA 2718
QY      2619 LLBKDLQIEQEEELKRDLLKKEKDLMLAQ-----ATAV 2653
Db      2719 STRHLETRVQK-----VQVKEEPSAVKFTQTSGETTDADKEPAGEDKGIKALKESAKQT 2773
QY      2654 AAPCPVPVPLPAPPAPPPPP-----PPFGVQHTGLLSTPTLPFVASQKRREBEKD 2705
Db      2774 PAPAASVTGSRPRAPRESAQALDLGPKDPAAGHTEESMTDDTKTKIPCKSSPELED 2833
QY      2706 SSSKSKKKMISTWETKDKTKLYCICKTPYDESKFYICDRCQWYHRCVGIQSEA 2765
Db      2834 TATSKRRPRTRAKRQVEKKEE--LLAVGK-----LTQTS 2866
QY      2766 ELIDEXVCPQOSTEDAMTTLPTLTKEDYGLKRLVLSLQAHKMWPFLPEVDPN-DAPD 2824
Db      2867 E-----THTDKPEVGEKGKTKA-----FKQPARNRNDAED 2897
QY      2825 YGVVKEPM-----DLATMEERVQR-RYYEKLTFEVAD 2856
Db      2898 VIGSRQPRAPKEKAQPLEDLASFQELSQTPPGHTEELANGAAD 2940

RESULT 7
ANK2_HUMAN
ID   ANK2_HUMAN          STANDARD;          PRT;   3924 AA.
AC   Q01484; Q01485;
DT   01-APR-1993 (Rel. 25, Created)
DI   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).
GN   ANK2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC   TISSUE=Brain stem;
RX   MEDLINE=91302466; PubMed=1830053;
RA   Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT   "Isolation and characterization of cDNAs encoding human brain
RL   ankyrins reveal a family of alternatively spliced genes.";
RL   J. Cell Biol. 114:241-253(1991).
RN   [2]
RP   REVISIONS.
RA   Carpenter S.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain stem;
RX   MEDLINE=94075409; PubMed=8253844;

```

```

RA      Chan W., Kordeli E., Bennett V.;
RT      "440-kD ankyrin: structure of the major developmentally regulated
RL      domain and selective localization in unmyelinated axons.";
RL      J. Cell Biol. 123:1463-1473(1993).
RN      [4]
RP      SEQUENCE OF 463-495 FROM N.A.
RX      MEDLINE=92009921; PubMed=1833308;
RA      Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
RA      Lux S.E., Ward D.C., Forget B.G.;
RT      "Isolation and chromosomal localization of a novel nonerythroid
RL      ankyrin gene.";
RL      Genomics 10:858-866(1991).
CC      -1- FUNCTION: Attach integral membrane proteins to cytoskeletal
CC      elements. Also bind to cytoskeletal proteins.
CC      -1- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=3;
CC      Name=1;
CC      IsoId=Q01484-1; Sequence=Displayed;
CC      Name=2;
CC      IsoId=Q01484-2; Sequence=VSP_000267, VSP_000268;
CC      Name=3;
CC      IsoId=Q01484-3; Sequence=VSP_000268;
CC      -1- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CC      CELLS THROUGHOUT THE BRAIN.
CC      -1- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
CC      AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
CC      AND FUNCTION (POTENTIAL).
CC      -1- SIMILARITY: Contains 23 ANK repeats.
CC      -1- SIMILARITY: Contains 1 death domain.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X56957; CAA40278.1; -
DR      EMBL; X56958; CAA40279.2; -
DR      EMBL; Z26634; CAB42644.1; -
DR      EMBL; M37123; AAA62828.1; -
DR      PIR; S37431; S37431.
DR      HSSP; P42771; IDC2.
DR      Genew; HGNC:493; ANK2.
DR      MIM; 106410; -
DR      InterPro; IPR002110; ANK.
DR      InterPro; IPR000488; Death.
DR      InterPro; IPR000906; Z05.
DR      Pfam; PF00023; ank; 23.
DR      Pfam; PF00531; death; 1.
DR      Pfam; PF00791; ZU5; 1.
DR      PRINTS; PR01415; ANKYRIN.
DR      SMART; SM00248; ANK; 22.
DR      SMART; SM00005; DEATH; 1.
DR      SMART; SM00218; ZU5; 1.
DR      PROSITE; PS50088; ANK_REPEAT; 20.
DR      PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR      PROSITE; PS50017; DEATH_DOMAIN; 1.
KW      Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW      Phosphorylation.
FT      REPEAT 63 92 ANK 1.
FT      REPEAT 96 125 ANK 2.
FT      REPEAT 129 158 ANK 3.
FT      REPEAT 162 191 ANK 4.
FT      REPEAT 193 220 ANK 5.
FT      REPEAT 232 261 ANK 6.
FT      REPEAT 265 294 ANK 7.
FT      REPEAT 327 360 ANK 8.
FT      REPEAT 331 360 ANK 9.
FT      REPEAT 364 393 ANK 10.
FT      REPEAT 397 426 ANK 11.
FT      REPEAT 430 459 ANK 12.

```


RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEvan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wozley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield J.S.N., Krzyzinski M.I., Skalska U., Smalus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RT proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5].
RP SEQUENCE OF 4405-5147 FROM N.A.
RA Kalicki J., Elliott G.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9Y6V0-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
CC VSP_003926, VSP_003927;
CC Note=No experimental confirmation available;
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 PDZ/DRH domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y19188; CAB60727.1; -;
CC EMBL; AC004903; AAD20936.1; -;
CC EMBL; AC004886; AAD21789.1; -;
CC EMBL; AB011131; BAA25485.1; -;
CC EMBL; BC001304; AAB01304.1; -;
CC EMBL; AC004082; AAB97937.1; -;
CC PIR; T00634; T00634.
CC HSSP; P04410; 1A25.
CC Genew; HGNC:13406; PCLO.
CC MIM; 604918; -;
CC GO; GO:0005856; C:cytoskeleton; NAS.
CC GO; GO:0045202; C:synaptic junction; ISS.
CC GO; GO:0005509; F:calcium ion binding activity; ISS.
CC GO; GO:0005544; F:calcium-dependent phospholipid binding acti. .; ISS.
CC GO; GO:0005522; F:profilin-binding activity; ISS.
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
CC GO; GO:0016080; P:synaptic vesicle targeting; ISS.
CC InterPro; IPR000008; C2.
CC InterPro; IPR001565; Synaptotagmin.
CC PRINTS; PR00360; C2DOMAIN.
CC PRINTS; PR00399; SYNAPTOTAGMN.
CC SMART; SM00239; C2; 2
CC PROSITE; PS00499; C2_DOMAIN_1; 1.

DR PROSITE; PS50004; C2_DOMAIN_2; 2.
KW Calcium/phospholipid-binding; zinc; Metal-binding; Zinc-finger;
KW Repeat; Alternative splicing.
FT NON_TER 1 465
FT DOMAIN 1 465
FT 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT P-A-K-P-Q-P-Q-P-X.
FT C4-TYPE (POTENTIAL).
FT C4-TYPE (POTENTIAL).
FT POLY-PRO.
FT PDZ
FT C2 DOMAIN 1.
FT C2 DOMAIN 2.
FT S -> SGNGLGIRIVGKEIFGHSGEIGAYIAKILPGGSAE
FT QTKLMEG (in isoform 2).
FT /FTid=VSP_003923.
FT K -> KPTDGGKVVSHPTGETQ (in isoform 2).
FT /FTid=VSP_003924.
FT G -> GOVMVONAS (in isoform 2).
FT /FTid=VSP_003925.
FT TAHKS -> SKRRK (in isoform 2).
FT /FTid=VSP_003926.
FT Missing (in isoform 2).
FT /FTid=VSP_003927.
FT SQ SEQUENCE 5147 AA; 563537 MW; CDS5D84990498CD3C CRC64;
Query Match 2.9%; Score 438.5; DB 1; Length 5147;
Best Local Similarity 18.2%; Pred. No. 5.6e-08;
Matches 554; Conservative 379; Mismatches 938; Indels 1171; Gaps 135;
Qy 275 LECVKPL--BEVP-----EDWQCEVCVAHKVPGVTCVAIEIKNKKYIRHEPTGY 324
Db 1076 LEVKETLSMEKIPPMVTDDQKESKLE-----KDKSALQEKKLPE----- 1119
Qy 325 DRSRRKYWFLNRLIIEEDTENENKKIWIYSTVKQLAELIDCLDKDYWEALCKILEE 384
Db 1120 -----EKKLPEEEKIRSEKK-----PLLEE- 1141
Qy 385 REEIRHRMDITEDLTNKGNSKFLAAANEIIELEIRAKKGDIDNVKSPETEKKRNE- 443
Db 1142 -----KKPTPEDKLLPEAKTSAPEE---QKHL--LKSQVOIAEEKLEG 1181
Qy 444 -----TENDSKDAKNREEFED---QSLEKSDDKT-----PDDPEQCK 480
Db 1182 RVAPKTVOEGKQPQTKMEGLSPGTPQSLPKD-DKTKTKIKEQPQPPCTAKPDQEKEDK 1240
Qy 481 SE-----EPTEVGDKGNS---VSANLGNTT-----NATSEETSPSE 514
Db 1241 SDTSSSQPKSPQGLSDTGYSSDGISSILGIPSLIPTDEKDIKGLKDSFSQESSPSS 1300
Qy 515 G-----RSPVGCILSETPDSSNMAEKVASELPQDV-PEEPNKTCESSNTSATTTSPN 567
Db 1301 PSDLAKLESTVLSILEAQASTLADKESEKKTQPEVSPQEP-----KDQEKTSLET 1353
Qy 568 LENSNSSELNSSESASAKAADDPENGERSHTPTVSTOEIVGDTSEK----STGELSE 623
Db 1354 LEITISEETIKESQERKDTFKDSDQDIPSKDHKEKSEFVDITTRREPYDVSSESE 1413
Qy 624 SPGAGKGASGSTRITITLRNP---DSKLSQLKSOQVAAAHANKLFKEGKVLVNVNSQ 680
Db 1414 SE-----NSPVQKRRRTSVGSSSDSEYKQEDSQSGSGEEDFIRKQ-----LIISAD 1461
Qy 681 EISRLSTKKEIVKIMGNINNVFKLGQEGKYRYVHYNQYSTNSPALNKHQHRDHRRLAH 740
Db 1462 EDAGSSEDEDFI-----RNQLKEIS--SSTESQKKE----- 1490
Qy 741 KFLTLPAGEFKWNGSVHGSKVLITLTLTTQLENNIPSPFFPNWASHRANIKAVQM 800
Db 1491 -----ETKGGKITAGK----- 1502
Qy 801 CSKPREFALALALECAVKPVMLPIWREFLGHTRLMHTSIREEKEKVKKEKKEE 860
Db 1503 -----HRLTRKSTSIDE-DAGRRHSHWDEDD 1529


```
Db 3265 SOGWTVSGDVGTOYI-----APP-----GILSTVSEIPLTDVVVKEEKQPKKR 3308
QY 2705 DSSSKKKKKMTSTSKETKDKTLKYLCKIPYDESKFYIGCDRCNNYHGRVCVILQSE 2764
Db 3309 SSGAKVRGQ-----YDDMGENTWDDP-----RSF 3332
QY 2765 ABLIDEYVCQCQSTEDAMTILPTLTERKDEGLKR-----VLSRLQAHKMAWPFLEVPDP 2819
Db 3333 KKVDSGV-----QTDD-----EDATRSYVSRRTKKSVDTSVQ-----TDD 3371
QY 2820 NDAPDYGVYKPEMDLATWEERQVR--RYEKKLTFEVDATMK 2859
Db 3372 EDQDEW-----DMPTRSRKARVKGYSMTS--ADTKK 3403

RESULT 9
BAZA_MOUSE
ID BAZA_MOUSE STANDARD; PRT; 1850 AA.
AC Q91YE5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bromodomain adjacent to zinc finger domain 2A (Transcription
DE termination factor-1 interacting protein 5) (TTF-1 interacting protein
DE 5) (Tip5).
GN BAZ2A Or Tip5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423468; PubMed=11532953;
RA Strohn R., Nemeth A., Jansa P., Hoffmann-Rohrer U., Santoro R.,
RA Laengst G., Grummt I.;
RT "NoRC-a novel member of mammalian ISWI-containing chromatin remodeling
RT machines.";
RL EMBO J. 20:4892-4900(2001).
CC -!- FUNCTION: May play a role in transcriptional regulation
CC interacting with ISWI. May serve a specific role in maintaining or
CC altering the chromatin structure of the rDNA locus.
CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NoRC
CC (nucleolar remodeling complex).
CC -!- SUBCELLULAR LOCATION: Nuclear. Co-localizes with the basal RNA
CC polymerase I transcription factor UBF in the nucleolus.
CC -!- SIMILARITY: BELONGS TO THE WAL FAMILY.
CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
CC -!- SIMILARITY: Contains 1 methyl-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ309544; CAC69992.1; -.
CC MGD; MGI:2151152; Baza2a.
CC DR GO: 0005731; C:nucleolus organizer complex; ISS.
CC DR GO: 0003677; F:DNA binding activity; ISS.
CC DR GO: 00030528; F:transcription regulator activity; ISS.
CC DR GO: 0006338; P:chromatin modeling; ISS.
CC DR GO: 0006355; P:regulation of transcription, DNA-dependent; ISS.
CC InterPro: IPR006637; AT_hook.
CC InterPro: IPR001487; Bromodomain.
CC InterPro: IPR004022; DDT_dom.
CC InterPro: IPR001739; Methyl-CpG_bind.
CC InterPro: IPR001965; Znf_PHD.
```

```
DR Pfam; PF02178; AT_hook; 4.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 4.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS00614; BROMODOMAIN_2; 1.
DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;
KW Nuclear protein; Repeat; DNA-binding.
FT DOMAIN 505 557 MBD.
FT DNA_BIND 602 614 A.T HOOK 1.
FT DNA_BIND 623 635 A.T HOOK 2.
FT DOMAIN 800 862 DDT.
FT DNA_BIND 1137 1149 A.T HOOK 3.
FT DNA_BIND 1351 1363 A.T HOOK 4.
FT ZN_FING 1623 1673 PHD-TYPE.
FT DOMAIN 1755 1825 BROMODOMAIN.
FT DOMAIN 613 738 LYS-RICH.
FT DOMAIN 647 774 COILED COIL (POTENTIAL).
FT DOMAIN 1006 1012 POLY-GLU.
SQ SEQUENCE 1850 AA; 205597 MW; E5DD4FED1D46DAE1 CRC64;

Query Match 2.8%; Score 422.5; DB 1; Length 1850;
Best Local Similarity 18.1%; Pred. No. 6.6e-08;
Matches 390; Conservative 295; Mismatches 772; Indels 693; Gaps 82;

QY 1082 STNSKNLSESPVI-----TKAKGQSDSMRQSQSNA 1115
Db 61 SANPCNNLKDPDLLSQFPGQYPLNGNRPSPSHNTNLRAGSEFWANGTQSPMG 120
QY 1116 NNDQPEDLIQGCQSDSSVLNRSDPSHTTNKLYPKDRVLDDVSIKSPETKCPKQNSIEND 1175
Db 121 LNFDSQELYDSDPDQNFVMPNGPPSFFTS---POTSPMLGSSI---QTFAPSDQ-VSSD 173
QY 1176 I-----EEKVSDLASRGQETPKSKTGNDFFIDSKIASADDITGLCKNKKPLIOEE 1228
Db 174 IHPDFAAREKELTSTVVAENG-----TGLVGSLEEEQELKWC 211
QY 1229 SDTIIVSSSKSALHSSVPKSTNDRDATPLSRAMDFEGLKGCDSSENSTLENSSDTVSIQDS 1288
Db 212 GYNGSVSVSESLHQEVSVLVPD---PTVSCLD-----DPSHLPDQ 248
QY 1289 SEEDMIVQNSNESISEQFTTREQDVEVLEPLKCELVSGESTG-----NCDRLPVKGT 1341
Db 249 LEDTPTLS-----EDSLEPFDLSAAEPVSGSLYGIDDAELMGAEDKLPLEG- 294
QY 1342 EANGKKPSQOKLEERPYNKCSQDIKLNKNTTDKKNENRESEKKGORTSTFQINGKNKP 1401
Db 295 -----NPVISALD----- 302
QY 1402 KIYLGECLEKISESRVSGNVEPKVNNIN--KIIPENDIKSLTVKESAIRFFINGDVIM 1459
Db 303 -----C-----PALSNANAFSLADDSDTSASIFVSPSPVLGSEVL 340
QY 1460 ED--FERNSETKSHLLSSDAEGNYRSLTFLTRESKSTQ-----TTTASACP 1510
Db 341 QDNSEGLNCSQDSQEIEIETQSS--NFQRLP-TEFAPDQPPSTQLHPAVSPFASPAALUT 397
QY 1511 ESNVNOVEDMEIETSEVKKVTSSPTSEESNLSNDFIDENGLPINKNENVGESKRKT 1570
Db 398 ASAEISPA-----VSPVASSVP----- 415
QY 1571 VITEVYTMSTVATESKTVKVEKDGKQTVVSTENCASKSTVTTTTTTVKLSTSTGGS 1630
Db 416 --PEVFAVSPASSPALPAISLE-----ASMTTPVTSPQSGSPSPAAAA 457
```



```
Db 1884 -----EQNG-----QKIRLREQIYDDPMQKITD----- 1914
QY 1379 NRESEKKGQRTSTFOINGKNPKIYLKGECLKEISRSVVGNEVPKVNINKIIPEND 1438
Db 1915 -----LOKEFEY----- 1933
QY 1439 IKSULTVKESAIRPFGDVIWEDFNERNSETSKSHLLSSDAEGBNYSRL--ETLPSTKE 1496
Db 1934 I-----VSSYIIPESHVIDLGSMTSTSEKK--LL--DADAAYEELMKROOMQVTDG 1984
QY 1497 SDSQTTPSASCPSNSVQVEDMEIETS-----EVKVTSSPITSEESNLSNDFIDE 1551
Db 1985 SSLIQTMTGDDMAESTLDFDRVQDASLTSSILSGASLTDSTSSATSLIPDKVIQHPSTE 2044
QY 1552 NGPINKNEVNGESKKKTVITEVMTSVATESKVIKVEKDKQTVVSSSTENCASKT 1611
Db 2045 E-----FEDEVYTDYTRIEQIIIAHESLITYSEPSATSVPPSDTSLTSSSVCTTD 2100
QY 1612 VTTTTVTVKLSTPSTGSGVDIISVKEQSKTVVT-----TVTDSL 1652
Db 2101 SSSPVTLDSLTVYT--EPADVIKFKDSEISTYFPGSVIDYPEDIGVSLDRTIPES 2159
QY 1653 TTTGGTLVTSM-----TVSKEYSTRDKVKLMKFSRKRKTRSGTALPSYRKFVTKSKKS-- 1706
Db 2160 RTNADQIMISFPGIAPSITESATKPE-----RQADTISTDLPISEKELIKGKETG 2212
QY 1707 -IFVLPNDLKLARKGGIREVYFNNAKPAIDWIPSPRPTFGITWYRLQTVKSLA 1765
Db 2213 DGIILEVLDAYKDKRESEAEKLSL--PETGLAPTSPSQT-----KEQGPSHSVS 2263
QY 1766 GVSMLRLWLASLRWDDMAKVPVGGG--STRTETSETIITTEIKRDRDVPYGRFE 1822
Db 2264 G-----EISGQKPYRSPGSLPVSTHPSKSHPPFRSSSLDISAQPPP----- 2307
QY 1823 YCIRKIICPIGVPETPKETPQKGLRSSALRKPETPKQ-----TGPVLIETWA 1875
Db 2308 -----PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 2356
QY 1876 EELELEWEIRAFARVEKEKAQVEQO-----AKRLEQQKPTVIATSTTSPT----- 1923
Db 2357 -----AHADAIPTEVATAARRSGLPATKICAAAAPPVPPKPSIPTGLVETHRP 2406
QY 1924 SSTTSTISPAQKVMVAPISGVTGKMWLTTKVGSPTVTFQONKFNHQTATWVKQGQ 1983
Db 2407 EASKPPIAPKPAVEIPVTIQTQTDTR--CPKPTGLPLTNSMSLNL----- 2449
QY 1984 SNSGVVQVQKVLGIIPSTGTSTSQTFSTQPR-----TATVTIRPNTSGSGGTTNS 2036
Db 2450 -----VTSADYK-----LPSPTSLPSHNSKSPRYSKSLMETYVVIILPSEPPTDSSAA 2501
QY 2037 QVITGPIRPGMTVIRTPLOQSTPLGKAILRTPVMQPGAPQVMTQI--IRGQPVSTAVSA 2095
Db 2502 QAITSS-----WPLGSPPKDLVSLVTFVS--VPPMTSTEIPSAQPTLVTSGA 2547
QY 2096 PNTVSTPG--QKSLTATSTSNLOSASOPPRQOQOVKLTMAQLTQLTQGHGNGGLTV 2154
Db 2548 LGTFSTVPAVTAFLQVTFVTSLTQFLPAEASKSEVSAVSAVSP-----RSVSI 2599
QY 2155 VI-----QGOQGTGOLQIIPQGVTVLPGQOL----- 2183
Db 2600 PIPPEPLADLRHOKENGKPLIGDAIDLTIPKSEKVKTEKMDLSASAMDKVKTOTAN 2659
QY 2184 -----MQAMPNGTVQRFLETPLATATTASTTTTSTTAAAGTEQROSKLSPOM 2234
Db 2660 EYVRRQISAVQPSIINLSAASSLGTPTVMTDSKTVAVVTCT-DTIYTTGTE-----SQV 2712
QY 2235 QVHQDKTLPPAASSVGPAKAPQTAQPSARPOQTPQOSPAQ-----PEVQ----- 2281
Db 2713 GIEHAVTSPQLTTS---RHTELQYKRPSSQAFPMIRDEAPINLSLGPSTQAVTLAVTKP 2769
QY 2282 -TOPEV-----QTQTTVSSHV--PSEAQPTHQSSKPVAAQSQPSNV-----QQQSPV 2328
Db 2770 VTPVPGVINGWTDTSIGITDGEVVVDLSTSKSHRTVTVMTDESTSNVTKIIEDEKPV 2829
```

RESULT 11

```
TRX_DROVI
ID TRX_DROVI STANDARD; PRT; 3828 AA.
AC Q24742; 2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trithorax protein.
GN TRX.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7244;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96100387; PubMed=8555104;
RA Tillib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
CC -!- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
CC WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -!- SIMILARITY: Contains 1 SET domain.
CC -!- SIMILARITY: Contains 5 PHD-type zinc fingers.
CC -!- SIMILARITY: Contains 1 post-SET domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```


Db 1632 --ELLSEQPPWFONETKACTDALEDNMFSCGVEELKESPTTYAEHHTASQAPRTGLLDI 1689
QY 1153 VLDVVSIRSPETKCPKONSIENDIEBKVSDSLARGOEPKSKTK-----GNDFIDDSKL 1207
Db 1690 PLDDVDDLQ---GC-----AVKRLDRVCLFCRKSGEGLSGEARILLYCHDCWVH--- 1738
QY 1208 ASADDIGTLCKNKKPLIQEESTIYSSSKSALHSSPKS-----TNDROT----- 1254
Db 1739 -----INCAMSAEVEFEID-----GSLQNVHSAVARGMKICTVCGNRGATVGCNVK 1786
QY 1255 -----PLSRAMD-----FEGKLGCDSSESNTLE--NSDTSVSIODSSEEDMIVONSNE 1301
Db 1787 SCGEHYHCARTIDCAFLTDKSMYCPAHARNALKANGSPSYIENFEV-----SRPV 1840
QY 1302 ISQFTRRODVEVLEPLKCEL-----VSGESTGNCEDR-----LPVKGTBA-----NG 1345
Db 1841 YVELERKRK---KLIIPAKVQFHIGSAVRQLGSIVPRESDSPEAIVPINFCLCSRLYWSS 1897
QY 1346 KRPSQOKKLEERP-----NKCSDQIKL---KNTDKKNENRESEKK----- 1386
Db 1898 KEP---WKIVEYTVRTIONSYSSTLTLDAGRNFYDHTNPNCSLVOLGLAQIARHSSL 1954
QY 1387 QRTSTEQINGKONKPIYKGECLKEISESRVSVGNVEPKVNNIKIIPENDIKSLTVE 1446
Db 1955 ARSDLLDTWAE--FPNSYVPAD---ENTEE-----EPQ-QNADLLPPE--IKD-AIFE 1999
QY 1447 SAIRPFING-----DVIM-EDFNERN-----SSEKSHLLSSSDAENGRDSLETLPSTKE 1496
Db 2000 DLPHELLDGLMIDIFWYEDLGKTELFAMSEKQKDTTATSQAGG--ASVVICDETRN 2057
QY 1497 SDS--TOTTPSASCPSNSVNVQ-----EDMEIETSEV-KVNTSSP-----I 1536
Db 2058 SNLSNKLHVLNSCCTASNPDVDAMLCAARSSSOEKECGVLKTKDTAPTSPWPKLDGGSV 2117
QY 1537 TSEESNLNDFIDENGLPINKNENVNGESKRTVITEVTMTSTVATESKTVIKVEKGD 1596
Db 2118 AAFKRRLSKNTAE--GVLLSLNQ-----RSKK-----EMATVAGITR-- 2153
QY 1597 KQNVSSSTENCAKSTVTTTTTKLSTPSTGSDVIISVKEQSK-----TVVTTT-VT 1649
Db 2154 ROSVCGSSELPAEGSATMTKSP-----TWSAAKCLFEKNESREEPAKLTIQMOMGDV 2206
QY 1650 DSLTT-----TGGTLVTSMTVK-----EYSTRDKV-----LMKSRPKKTRS 1688
Db 2207 DSITEYRIIGSDGNLSHTAQTGOVKERCOCQYRNYDFSQRHLGSCPEMSTSESESTAT 2266
QY 1689 GTALPSYRKFVTKSKKSIFFVLPNDLKK-----LARKGIREVPYFYNKAPALDIW 1741
Db 2267 GTAQLSAESL-----NELQKALAAATLSNTGGL-----NY-----LQTSF 2302
QY 1742 PYPSPRPT---FGITWRYRLQTVKSLAGVSL-----MLRLWLASL 1778
Db 2303 POVQNLATLQFGVQVQLOGLOTLIQ-LQPSQSLGNFPLSQPNAAQATSNQNDVLQLYANSL 2361
QY 1779 RWDMAAKYPPGGSGSTRSETSEITEITTEILKRRDVGPGYRFEYCIKILCPIGVPETP 1838
Db 2362 Q---NLAAANL--GGGFLTQPTMSTQAPQLI--ALSTNPDG-----TQ 2398
QY 1839 KETPTPQRKGLRSSALR---PKRPETPKOTGPVITETWAELELELWEIRAFARVEKEK 1895
Db 2399 OFIQLPQNGATTOLOTAAPLRCNATYOT---LOATNSDKKIVLLFLEA-GDPLQEVV 2453
QY 1896 AQAVEQ-----QAKRLEQQKPTVIATSTPTSSTSTTSPISPAKV 1936
Db 2454 TOAAQATAAAHOKLKGSHGVKPIQAKLQGOQOQOQHQQHQQHQQOQOQOQOQOQOQ 2513
QY 1937 MWAPISGVTGKWLVTTKVSGPATVTFQOKNHFOTFATVWQCSNSGVVVOQKVL 1996
Db 2514 QOTPTIVAOHGGTTOLLGONLQIP-QLLFQSNQAQ-PQOQOQLLPQTAQN----- 2561
QY 1997 GIIPSTGTSQO-----TFTSFOPRTATVIRPNTSGSGGT----- 2033
Db 2562 -IISFVYTGGSQONQLOYSIPTINDFKPQQTSTPTFLTAPCGGATFLOTDAAGNMLT 2620
QY 2034 -----SNSQVITGPQIRPGMTVIRTPLOQSTL-----GKAIIRTPVMVQP----- 2073
Db 2621 TAPANSGLQMLTG-QLOTOPOVIGTLIQPTQLQTTGADGTQTATATAQOPLILGATGGGT 2679
QY 2074 -----GAPOQVMQTIIRGPV-----STAVSAPNTVSS-T-PG---QKSLTSATST 2114
Db 2680 TGLEFATAPOVILAT---QPMYIYGLTIVONTVMSSQQFVSTAMPQVLSQNSFSFATTT 2735
QY 2115 SNIQSSASOP---PRP-----QGOVKLTMAQLTQLTQGHGGNGLT 2153
Db 2736 QVFQASKIEPIVDLPAGYVVLNNAVDSAGNTSWLQSQTOATDDATAQLLQ-NAGFOFQT 2794
QY 2154 VTIQGGQTTGQLOLIPQGV-----VLPQPGO-OLMOAA 2187
Db 2795 TPTTSTQMTSTDYAPLVVTAQVPPVAQMKRNTNANKSPISVLSKVOPQOQSVVNVK 2854
QY 2188 MPNGTV-----QRFLFTPLATATTASTTTTT-----VSTIAG--T 2222
Db 2855 LPTNVITQO 2914
QY 2223 GEORQKSLSPOMOVHODKT--LPPAQSSSVGP--AKAQPOTAPSPARPQOTOPQSPA-- 2276
Db 2915 GSTCGAPPSTASKPLOKKNLIRPIHKVEVKPKIMQAPKLATSAASHQHQOQOQSPAAI 2974
QY 2277 -----QPEVQTPQEVQTTQTVSSHVPSSEAPTHAQSSKRPQVAAASQPP 2319
Db 2975 NOVAKVALLQORLAPAPQO-QQEPQEEQHLHQOQOQOQOQOQOQOQOQOQOQOQOQO 3033
QY 2320 SNVOGSPV-----RVQSPQTRIRPSTPQSLSPQOQSOVQTTSTPITP----- 2363
Db 3034 L-LRAQOPIISIVNTAPEQAATQFVIRPALQAQAPITQLOEQSQOQOQOQOQOQOQOQO 3092
QY 2364 --IOPHTSLQIPSQ-GOPOSOPOVOSSSTOPLSSGOTLNVQSVSSPSR-----PQLQIQO 2414
Db 3093 ARLORYASNSLPTNVNPLQOQRCASANNSSNNTQONSTIINSRPTNRVLPQOQOQO 3152
QY 2415 PQP-----QVIATP-----QLQOQOQOVLQIQSOQV---AQIQAQOQSGVPOQIKQLPT-QI 2462
Db 3153 PTPLSNDVVVQSPPTPKPIEPPVAGASTOKPVKCYAQLEQKSPGYETELKTNITLDNL 3212
QY 2463 QQSSAVQTHQIQ-----NVVTQAAASVQOLQR-----VQOLRDQOQKQOQO 2505
Db 3213 EOTNITTQLOQO 3272
QY 2506 IEINVTNPKLLIKVEILQOVVVKHNAVIE-----HLKQKKS-----MT 2545
Db 3273 MEMVVDNGFQLTSN---ESCLLEKHGFNVPAVPMDETHYASKNKGSGGAAEGIGQVD 3328
QY 2546 PAEREN-----QRMIVCN-----QVMKYILDKIDKEE---KOAARKKR 2582
Db 3329 DABEEDDDDDDFSLKMATSNACNDHMSDSSEPAVKEKISKILNDLTNDGSDSTATATTV 3388
QY 2583 EESVEQKRSKON--ATKLSALLPKHKEQLRA-----EILKRALLDKLOLIE 2627
Db 3389 EASAGYQOQVVEDVLATTAAGSVSTDDTTTATAEVAASYNEMAEAEHELQOLQOAG 3448
QY 2628 VQBELKRLKIKKEKDLMLQALQATAVAACPPTVPLPAPPAPPPPPPPPPVQHTGLLS 2687
Db 3449 VELDLK-----KPKLDVPOQOQDTPV-PNVVPTAAAPQOQPPMRDPKK-----IS 3492
QY 2688 TPTPLVASQKRREKREKSDSSKSKKKMISTTSKETKDKTKLYCICTKPYDES----- 2740
Db 3493 GPHLLYEIQSDEGFTYKSSSIAIEWEKVFQVARRAHG-----LTPLPEGLADMSG 3546
QY 2741 -----KFYI-----GCDRCQWY-----HGRCVG-----ILOSE 2764
Db 3547 VOMIGLKTNALYIELQPLGVEKVCYKPYTKYHKNRGNVSTAAGGHARTAGSNPAALAG 3606
QY 2765 AE-LIID-----EYVCPQOCQSTEDAMTTLTLETKDYBGLKRLVLSLQAHKMAWP 2812
Db 3607 AESLIDYGSQOBELOENAYECARCE-----PYVSRSEYDMFSLASRRHKQPIQV 3656


```
QY 990 LDSDDPK-----CKEPEVDDDDMKTESHVNCQESSQVDVNVNVEGFLHRTSYKKK----- 1041
DQ 998 QDGLSDIAEFLKQSDSEDDSKQKKGTEKK-----PSDFKKVYIKM 1005
QY 1042 -----TKSKLDGLLER-----RIKQFTLEEKORLEKIKLEGGIK----- 1076
DQ 1006 EQQYESSDGTETKLEPEREICHFPKGIQIKNGTTGDEKSKKIKDKTSKKKDELSDYAE 1065
QY 1077 ---GIGKTSNKNLSESVITKAKEGQ---SDSMRQEQSPNANNDQPEDLIQCSQSD 1131
DQ 1066 KSTGKGDSDSSDKSKNGAYGREKCKLLGKSRKQD-----CSSSD 1111
QY 1132 SSVLRMSDPSTHNTKLYPKDRVLDVSTRSPETCKPKONST-----NDIEEKVSDLA 1184
DQ 1112 TEKYSMKDGCNNS-----DKRLKRIELRRRLNLSKRTKEIQSGSSSSDAEESSEDNK 1166
QY 1185 SRGOEPTKSKTKGNDFIDDDSKLASADDIGTLCKNKKPLIQEESDTIVSSSKSALHSV 1244
DQ 1167 KKKQR-TSSKKK-----AVIVKEKK-----RNSLRST 1193
QY 1245 PKSTNDRDATPLSRAMDFEGKLGCDSESNSTLSESDTVSIQD-----SSEDMIVQN 1297
DQ 1194 KRQKQADITSSSSSDIE-----DDQNSIGEGSSDEQKIKPVTEINLVSSHTGFCQSS 1245
QY 1298 SNESISEQFRTREQDVE-----VLEPLKCELVGSESTGNCEDRLPVKGTGAN 1344
DQ 1246 GDEALSKSPVTVDDDDDDNDPENRIAKMLLEETKANLSSDED-GSSDDE-PBEGKKRT 1303
QY 1345 GKPSQOKKLEE-RPVNKCSDQ-----IKLKNNT-----DKNNENRE 1381
DQ 1304 GKQNEENPCDEAKNQVNSDSGSESKKPRYRHLRLHLTVSDGSGSEKKTYPKEH 1363
QY 1382 SEKKQRTSTEQINGKNKPKIYLLGCECLKBISRSVGVNVEKVNINIKIIPENDIKS 1441
DQ 1364 KEVKGRRNR-KVSESDSESDQESGVSESVES---EDEQRPTRSAKKALEENQRS 1418
QY 1442 LTVKESATRPINGVIMEDNERNNSSTKSHLLSSSDAEGNRYDSLTLPSTKES----- 1497
DQ 1419 YKOKKRRIRIKVQEDSSSE--NKSNEEEBEKEEEEEEEEEEDENDDSKSPGKG 1476
QY 1498 -----DSTOTTPSPASCPESNSVNOVEDMEIETSEVKV-----TSSPINS- 1538
DQ 1477 RKKIRKILKDKLRTETONALKEEERKRIAEEREREKREKREKREKREKREKREKREK 1536
QY 1539 -----EESNLSNDFIDENGLPINKNENVG-----ESKRT-----VITEVT 1576
DQ 1537 LVLDEDETEKPLQVQVHRNVWIKLPHQVGVQFMWQCCESVKKTKKSPGSGCILAHC 1596
QY 1577 TWTSTVATES--KIVIKVEKGDQTVV 1601
DQ 1597 GLGKTLQVYVFLHTVLLCDLDFSTAL 1623
```

RESULT 13

```
ANK3_HUMAN
ID ANK3_HUMAN STANDARD; PRT: 4377 AA.
AC Q12955;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ankyrin 3 (ANK-3) (Ankyrin G).
GN ANK3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain stem;
RX MEDLINE=95138209; PubMed=7836469;
RA Kordeli E., Lambert S., Bennett V.;
RT "AnkyrinG. A new ankyrin gene with neural-specific isoforms localized
at the axonal initial segment and node of Ranvier.";
```

```
RL J. Biol. Chem. 270:2352-2359(1995).
CC !- FUNCTION: Membrane-cytoskeleton linker.
CC !- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=1;
CC Comment-A number of isoforms are produced;
CC Name=1;
CC IsoId=Q12955-1; Sequence=Displayed;
CC !- TISSUE SPECIFICITY: Expressed in brain and other tissues.
CC !- SIMILARITY: Contains 23 ANK repeats.
CC !- SIMILARITY: Contains 1 death domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DQ EMBL: U13616; AAA64834.1;
DQ PIR: A55575; A55575.
DQ HSP: P55273; 1B18.
DQ Genew: HGNC:494; ANK3.
DQ MIM: 600465;
DQ GO: GO:0006605; P:protein targeting; NAS.
DQ InterPro: IPR002110; ANK.
DQ InterPro: IPR000488; Death.
DQ InterPro: IPR000906; ZU5.
DQ Pfam: PF00023; ank; 24.
DQ Pfam: PF00531; death; 1.
DQ Pfam: PF00791; ZU5; 1.
DQ SMART: SM00248; ANK; 21.
DQ SMART: SM00005; DEATH; 1.
DQ SMART: SM00218; ZU5; 1.
DQ PROSITE: PS50088; ANK_REPEAT; 21.
DQ PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DQ PROSITE: PS50017; DEATH_DOMAIN; 1.
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat.
FT REPEAT 73 102 ANK 1.
FT REPEAT 106 135 ANK 2.
FT REPEAT 139 168 ANK 3.
FT REPEAT 172 201 ANK 4.
FT REPEAT 203 230 ANK 5.
FT REPEAT 234 263 ANK 6.
FT REPEAT 267 296 ANK 7.
FT REPEAT 300 329 ANK 8.
FT REPEAT 333 362 ANK 9.
FT REPEAT 366 395 ANK 10.
FT REPEAT 399 428 ANK 11.
FT REPEAT 432 461 ANK 12.
FT REPEAT 465 494 ANK 13.
FT REPEAT 498 527 ANK 14.
FT REPEAT 531 560 ANK 15.
FT REPEAT 564 593 ANK 16.
FT REPEAT 597 626 ANK 17.
FT REPEAT 630 659 ANK 18.
FT REPEAT 663 692 ANK 19.
FT REPEAT 696 725 ANK 20.
FT REPEAT 729 758 ANK 21.
FT REPEAT 762 791 ANK 22.
FT REPEAT 795 825 ANK 23.
FT DOMAIN 1519 1898 SER-RICH.
FT DOMAIN 4090 4174 DEATH.
SQ SEQUENCE 4377 AA; 480399 MW; F42379E55768B684 CRC64;
```

Query Match 2.7%; Score 406; DB 1; Length 4377;
Best Local Similarity 17.6%; Pred. No. 6.5e-07;
Matches 558; Conservative 444; Mismatches 1118; Indels 1048; Gaps 126;

QY 17 QDSFDEDEMEDEDDSDSY-----PEEMEDDDDA----- 47
DB 1441 KETESDQDEIEKTDRRQSFASLAKRYSYLTPGMIERTGATSLPTTYKYKPFST 1500

```
QY 48 ----SYCT-----ESSFRSHSTYSSTGRKKP-----RVHRPSPILEEKDIPPLE 89
Db 1501 RPYOSWTAPITVPCPACSGFTLSSSSNTPSPASPLKSIWSTP-SPKSTLGASTTS 1559
QY 90 FPKSSEDLMPVNEHIMVIAIYELRNFVGTVLRLSPRFEFCAALYSQEOCTLMAEHV 149
Db 1560 SVKISDVASP-----IRSLRTMSSPIKTIVVOSPNYQVSSGTLARAPAVTATPLKG 1613
QY 150 VLLKAVIREEDTSNTTFG-----PADLKDSVNSTLYFIDGMTWPEVLRVCES 197
Db 1614 LASNSTFSRTPSVTTAGSLERSSITWTPASPKNINM-----1653
QY 198 DKEYHHVLPYQ-----EADYPYGVENKIKVLQFVDOFLTNTARELMSEGVQYDDH 253
Db 1654 ---YSSLPEFKSIITSAAPLISSPLKSVVSKRVDVVISSAKITMASSLPVKOMPGH 1710
QY 254 CRVCHKLGDL-----LCCETCSAVYHLECVKPLEEVPEDQCEVCAHKVPCVT 304
Db 1711 AEVALVNGSISPLXYASSSTLINGCKATATLQ-----EKISSATNSVSVVSAAT 1760
QY 305 DCVAEIQKN-----KPYIRHEPIGYDRSRK-----YWFNLRRLLIIEEDTENENEK 350
Db 1761 DTVEKVFSTTTAMPFSLRSVSAAPSAFQSLRTPSASALYTSLSGSSISATTSVTSSII 1820
QY 351 KIWTYST-----KVQLAELIDCLDKYWEAELKILEMREEIH--RHMDITEDLTNKAR 403
Db 1821 TVPYVSVVNLPEPALKLPDSNFTKSAALLSPIKTTLTETHPHQPHFSRTSPV--- 1876
QY 404 GSNKSFLLA-----AANEELIESIRAKGDDIDNVKS-----PEETE 438
Db 1877 -KSLFLAPSALKLSTPSSLSSSOEILKDVAMKEDLMRTAIILOTDVPEKPPQPELPK 1935
QY 439 KDKNETENDSKDAENREFEQS--LEKD--SDKTPDDOPEQGKSEEPTEVGDGKNSV 494
Db 1936 EGRIDDEPEFKIVEKVEDLVKVSLEILKDKVCVONK--GSPKSPKS-----DKGHP 1985
QY 495 SANLGDNNTNATSEETSPSEGRSPVGLSETPDSNMAEKVASELPQDVPEEPN-KTCE 553
Db 1986 E-----DOWIESSEI--REAR-----QAAASQSP-SLPERVQVRAKA 2022
QY 554 SSNTSATTSTIQPLENSSELSSNSESASAKAADPPENGERSHTP-VSIOEE----- 607
Db 2023 ASEKDYNLTVIDYLTNDIGSSLTNLKYFPEDAKKGEGQKRVLPALQAEHKLMP 2082
QY 608 --IVGDTSEKSTGELSESPGAGKAGSGSTRIITRLRNP-----SKLS-----QLK 652
Db 2083 PASMRTSTSEKELCKMADS-----FFGTDTI--LESPDDFSHQDDQKSPLSDSGFETR 2133
QY 653 SOQVAAAHEA-----NKLPE-----GKEVLV-----NSOGEISRLTKKEVI 692
Db 2134 SEKTPSAQSAETGPKPLFHEVPPIPVITETRETVVHVIRKSYDPSAGDVPQTQPEBFS 2193
QY 693 MKGNNVYFKLQEGKYRVYHNOYSTNSFALNKHQHREDHDKRHLAKHFKCLTTPAGEFKW 752
Db 2194 PKPS-PTFEMEL--EPKPTTSIKEKVAFQMKASSEEDHN--RVLS-----2235
QY 753 NGSVHGSKVLTISTRLTITQLENNIPSSFHPNASHRA-----NWKAVQMCKPR 805
Db 2236 ---KGMRVKEETHITTTTRVYHSPPG---EGASERIBETMSVHDIMKAFQSGRDP 2287
QY 806 EFALALAILCAKVPVNLPIWREFLGHTRLRM---TSTEREKEKVKKKEKQEBEET 862
Db 2288 KETAGLFEHKSAVSDV-----HKSAAETSQAHAEDKNQMPKLERIIE-- 2331
QY 863 MQQATWVKYTFPVKHVYKOKGEEYRVTYGGWSWISKTHVYRVFVKLPGLNTNRYKSL 922
Db 2332 -----VHIEKNQAEPTV-----LIIRTKKHPEKEMVYQKDL 2365
QY 923 -EGTKNNMDENMDESRRKRSRPPKIKIEPDSKEDVKGSDAAKAGADQONEMDISKITEK 981
Db 2366 SRGIDINLKDFLPEKHADFPCSEE-----OQOQEBELTAEESLPVLESRRVTPVSQE 2419
QY 982 KQO--DVKELLDSQDKPKC---EPEMEVDDO-----MKTESHVNCQESSQVDVNVVNSEGF 1032
```

```
Db 2420 EDSRPSSAQLISDDSYKTKLLSQHSIEYHDELSELGSESYRFAEKMLLSEKLDVS--- 2476
QY 1033 HLRTSYK---KTKSSKLDGLLR-RIKQFTLEBKORLEKIKLGGKIGTKGTSTNSK 1087
Db 2477 HSDTEESYTDIAGPPSELQGSDDKRSREKIATAPKEILSKYIKVDVSGVGVKS----- 2531
QY 1088 NLSESPVITKAKEGCQSDSMRQEQSPNANNDQPEDLIQGCQSQSDSVLRMSDPSHT--- 1143
Db 2532 -----KDEHFDKVTVLHYSGN---VSSPKHAMMR 2558
QY 1144 -----TNKLYPKDRVLDVDSIRSPETTKCPQONSIENDIEEKVSDIASRGQBP-TKS 1193
Db 2559 FTEDRLDRGREKLIYEDRV--DRTVKEAEELKTEVSQFFRDKTEKLDN---ELQSPKKA 2613
QY 1194 KTKGNDFPIDDSKLASADD--IGTFLICKNK-----PLIOESDPIVSS 1236
Db 2614 RPKNGKEYSSOSPSTSSPEKVELLTLLASNDENWKAROHGPDGQGFPAEKAPSLPSP 2673
QY 1237 KSALHSSVPKSTNDRDAT-----PLSRAMD-----FEGKLGDS 1270
Db 2674 EKWVLS---QOTEDSKSTVEAKGISQSQKAPDGPQSGFQKQSKLSSIRLAKFEQTHAKS 2730
QY 1271 ESNSTLENSDTVS-----IQDS-----SEEDMIVQNSNESISEQ-----FTR 1309
Db 2731 KMSQEDRKSDGSRIPVKKIOESKLPYVQFAREKQOKAIDLDPDESYSQKDFMVLTK 2790
QY 1310 BODYEVLE-----PLKCELVSGESTGNCEDRL 1336
Db 2791 DEHAQSNIEVYNDGSDNVKKQRTSEMSSKAMPDSFSEQAKDLACHITSDIATRGPDWK 2850
QY 1337 PVKTEANG--KPKSQKKLEERPYNKCSQDIKLTNTDKKNE--NRESEKKGORTSTF 1392
Db 2851 VFTWESSGATNNKSQEKLSHLVH-----DVRENHIGHPESKSDVKNEFM 2898
QY 1393 QINGKDNKPKIYLKGECLKEISESRVSVGNVEPKVNNINKIIPENDIISLTVKESAIRPF 1452
Db 2899 SVTERERK---LLANGSLSEIKEMTVKSPS-----KKVLYREYVVRK 2938
QY 1453 INGDVIMEDFNERNSETKSH-----LLSSDAEGNVRDSTLTPSTKESDSTQT 1502
Db 2939 DHPGLLDQPSRRESSAVSHIPRVADERRMLSNIPDGPCEQS--AFPHKLSQKLSQ 2996
QY 1503 TTPSASCPESNVQVEDMEIETSEVKVT-----SSPITSEE-----1540
Db 2997 SSMSKETVETQHFNSIEDEKVTYSEISKVSHQSVGLCPPLEETETSPTKSPDSLEFSP 3056
QY 1541 -ESNLNDFIDENGLPINKNENV-----NGESKRKTVITEVTTMTSVATESKTVIKVE 1593
Db 3057 GKESPSSDVFDHS--PIDGLEKLAPLAQTEGGKEIKLTPVYV---SFVQVGQYKEIQ 3110
QY 1594 KGDKQTVVSSPENCASKTVTVTTLSTPSTGGSVDIISVRBQSKVTVTVTVDLSLT 1653
Db 3111 QGGVKKIISQCKTVQETRGTFYTTROQKQPPSQGSPEDDTLEQVS-----FLD 3160
QY 1654 TTGGLTVTSMT-----VSKEYSTRDKVKLMKF-----SRPKTRSGTALPS 1694
Db 3161 SSGKSPLTPETPSSEESVYFTSKTPDLSLIAYIPGKPSPIPEVSESEEEQAKSTL-- 3218
QY 1695 YRKFTVSTKKSIFV--LPNDLLKLLARKGIREVPFVFNNAKPALD-----1739
Db 3219 ---KQTTVEETAVEREMPNDVSKDSNQRKNRNVAYIEFPPLPPDLADQIESDKHHY 3273
QY 1740 -----IWPVSP-----RPTFGITW 1754
Db 3274 LPEKEDMIEVNLQDEHDKYOLAEVPIRVQP-PSVPPPGADVSSDDESIIYQVPVKKY 3332
QY 1755 RYRLQTV-----KSLAGVSLMLRLWAS-----LRWDDMAAKVPPGGGSTRTE 1797
Db 3333 TFKLKEVDDEQEKPKASAEKASAKQELSGKDNFEGGLDLSPOEIAQNGNDQSI 3392
QY 1798 TSETIITTEITIKRRD-----VGPYGIRFE-----YCIRKII 1829
```


Db 244 TSSTSSSTSSSTAPATPTTTSCTKEKPTPTTTSCTKEKPTP-----PHDHTPCT 296
QY 1367 KLNNTDKNNENRESEKKQORTSTFQINGKDNKPKIYLKGECLKEISRVVSGNVEPK 1426
Db 297 KKKTTTSKCTKKTTPVPPTSSSTTE-----SSAPVPTPSSSTSSAPVTSSTTE-- 350
QY 1427 VNNIKIIPNDIKSLTVRESATRPINGVDIMEDNERNSETKSHLLSSDABGNRD 1486
Db 351 --SSAPV--TPSSSTSSAPVTS-----STSSAPVTSSTSSAP 394
QY 1487 SLETLPTSKESDS---TOTTPSAPCESNVNOVEDMETETSEVKKVTSSPTSE--- 1539
Db 395 VPTPSSSTSSAPVTSSTSSAPVTSSTSSAPV--TSSTSSAPVTSSTTES 453
QY 1540 -----EEN--LSNDFIDENGLPINKNENNGESKRKTIVTEVTMTST--VA 1583
Db 454 SSAPVPTPSSSTSSAPVTSSTSSAPVPTPSSSTSSAPVTSSTSSAPV 513
QY 1584 TESKTVKVEKDKQIVVSTENCASKSTVTTTTVTVKLSTPGSGVDIISVKEQSKTV 1643
Db 514 TPSSSTSSAPVPTPSSSTSSAPVTSSTSSAPVPTPSS-----STSSSTP 568
QY 1644 VTTVTDSLTGTGTLVMTVSKSKE--YSTRDKVKLMKFRPKKTRSGTALPSYKRVTK 1701
Db 569 VTSSTSSAPVPTPSSSTSSAPVPTPSSSTSSAPVPTPSSSTSSAPVTS 628
QY 1702 STKKSIFVPLNDLKLARKGGIREVPYFYNNAKPALDIWPPSPRPTFGITWYRLQTV 1761
Db 629 STSSSSA-----PVPTPSS-----TTE 647
QY 1762 KSLAGVSLMLRLWASLRWDDMAAKVPPGGSGTRTETSETIIT-----TEIKR 1811
Db 648 SSAPVPTP-----SSSTSSAPVPTPSSST--TESSAPVTSSTSSAPVTSSTTE 701
QY 1812 RDVGPGYRPEYCIKILICIGVP-----ETPKETPTQKGLRSSAL----- 1854
Db 702 SSAPVPTPSSSTSSAPVPTPSSSTSSAPVPTPSSSTSSAPVTSSTSS 761
QY 1855 -----RPKRPTPKOTGPVITIVWAEELWEIRAFERKEKAQAVEQQAKRLEQQ 1910
Db 762 APVPTPSSSTSSAPVPTPSSSTSSAPVPTPSSSTSSAPVPTPSSSNTSS 821
QY 1911 KP--TVIANTTS-----PTSSSTSTSPAKVMVAPISGVTGTCKMVLTKVGS-- 1959
Db 822 APSSTPSSSTSSVPTPSSSTSS-----APVSSSTSSAPVPTPSSSN 874
QY 1960 -----PATVTFQONKFNHOFATWVWKGOSNGVQVQKVLG-----IIPS----- 2001
Db 875 ITSAPSSIPFSSTTESFST-GTVVTPSSSKYPGSQTESVSTTETTVPTKTTTSVTT 933
QY 2002 -----STGTSQ--QTFTSFQPRATVTRPNTSGSGGTTNSQVITGPQIRPGM 2048
Db 934 PSTTTTITTVTCSTGNSAGETSGCSPKTVTTTV-PTTTTTSVTSSTTIT-----T 985
QY 2049 TVIRTPLQ---QSTLG---KALIRT-PVMVQCAPQVMQIIRGPVSTAVS-----AP 2096
Db 986 TVCSTGNSAGETSGCSPKTIITTVPCSTSPSETASESTTTPPTVTTVSTTVTTE 1045
QY 2097 NTVSSTPGOKSLTSATSTSNISQS---ASQPPRQOQOV-KLTMAOLTOLTQGHGNOGLT 2153
Db 1046 YSTSTKPGGE-INTTFVTKNIPTYLITTAFTSVTVTNFTTTTITTVCSGTGNSA-- 1102
QY 2154 VTIQGGQTTGQIQLIPQGVTVL-----PGPGQOLMQAAMPNGTVORFLETPLATATTAS 2209
Db 1103 -----GETTSGCS--PKVTVTVPCTSGTGEVTEA-----TTLVTTA--- 1138
QY 2210 TTTTTSVTAAGTGQRQSKLSQMVHQDKTLLPPAOSSVSGPAKAQPTQAQPSARPQ 2269
Db 1139 -VTTTVTTESSTGTSAGKTTTG---YTTKSPVTTVTTLAPS----- 1178
QY 2270 TQPSAPQVQVQPEVQVQTTTVSSHVPSAPQTHAOSKPKQVAAQSQPNVQGSQSPVR 2329
Db 1179 ----APVTATNAPVTITTTESCAATNAGETTSVCSAKTIIVSSASAGENTAPSAT--- 1231

QY 2330 VQSPSQTRIPRST--PSQLSPGQSQVQTT---TSQPIPIQPHTSLSQIPSGQSQSQPQV 2384
Db 1232 --TPVTTAI-PTTVITTESSVGTNSAGETTTGTTKSIPTTYITTL-IPGNGAKNYETV 1287
QY 2385 QSSTQTLSSQTLNOVSVSPSRPOLQIQPOPOVIAVPOLQOQVQLSQSQVQAQIQ 2444
Db 1288 ATAT-----NPISIKTSS-----QLATTASASSVAPV 1315
QY 2445 AOSGVPQIKLOLPIQIOIOSSAVOTHQIONV-VTVQAAS 2483
Db 1316 TSPS-----LTGPIQASGSAAVATISVPSISSTYOGAA 1348

RESULT 15

ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN MUC2 OR SMUC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R. Jr., Cuihane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC AGENTS AT MUCOSAL SURFACES.
CC -1- SUBUNIT: MULTIMERIC.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC -1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC VARIES AMONG DIFFERENT ALLELES.
CC -1- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC -1- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC -1- SIMILARITY: Contains 2 WVC domains.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2003, 16:10:05 ; Search time 56.2184 Seconds
(without alignments)
8207.607 Million cell updates/sec

Title: US-09-698-295-1
Perfect score: 14971
Sequence: 1 MVSEEEEDGDAETQDSE.....KLGFKASRNNKLOSTAS 2907

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	14971	100.0	2907	21	AA197452 Human transcriptio
2	14170	94.6	2781	21	AA157453 Human transcriptio
3	13735	91.7	2759	24	AA016418 Human nucleic acid
4	2695.5	18.0	557	22	AA894078 Human protein sequ
5	1488.5	9.9	976	22	AB86735 Drosophila melanog
6	1385	9.3	328	21	AB54364 Human pancreatic c
7	1030	6.9	191	24	ABU70846 Human adipocyte Se
8	982.5	6.6	238	20	AA104323 Human secreted pro
9	981	6.6	237	22	AAU19446 Human diagnostic a

10	789	5.3	149	23	AAU78845	SCAM/KRAB protein
11	781	5.2	246	22	AAU31944	Novel human secret
12	661.5	4.4	997	22	ABG62887	Drosophila melanog
13	621.5	4.2	2768	22	ABG68397	Drosophila melanog
14	613.5	4.1	5533	22	ABG65772	Drosophila melanog
15	613.5	4.1	5560	22	ABB71160	Drosophila melanog
16	567	3.8	6815	22	ABG66811	Drosophila melanog
17	562.5	3.8	3257	22	ABG67502	Drosophila melanog
18	561	3.7	3080	22	ABG64877	Drosophila melanog
19	551.5	3.7	2897	22	ABG58514	Drosophila melanog
20	503.5	3.4	4498	22	ABG58595	Drosophila melanog
21	491	3.3	2137	23	ABF39618	Drosophila melanog
22	474	3.2	3726	22	ABG66878	Staphylococcus epi
23	462	3.1	2586	22	ABG66878	Drosophila melanog
24	460	3.1	10431	24	ABU54861	Human CA125 amino
25	458.5	3.1	3266	21	AA842491	Human ORFX ORF2255
26	456.5	3.0	3664	24	ABR47592	Breast cancer asso
27	456	3.0	3021	24	ABP75877	Human secretory po
28	451.5	3.0	2283	24	ABP56876	Staphylococcus epi
29	449.5	3.0	3111	22	ABG60327	Drosophila melanog
30	448	3.0	3236	24	ABP96155	Human Ki-67 chromo
31	444	3.0	2977	22	ABG69480	Drosophila melanog
32	443	3.0	2261	24	ABU18914	Pathogen specific
33	443	3.0	2344	22	ABU37120	Staphylococcus aur
34	442	3.0	1870	24	ABJ19019	Pathogen specific
35	441.5	2.9	3696	23	ABP40235	Staphylococcus epi
36	439.5	2.9	1878	22	AAU40239	Human polypeptide
37	439.5	2.9	3256	21	AAU50976	Human cell cycle p
38	439.5	2.9	3256	24	ABU07489	Protein different
39	438	2.9	3256	23	ABG77188	Prostate adenocarc
40	435	2.9	1727	22	AAU95554	Human protein sequ
41	433	2.9	1795	22	ABG69806	Drosophila melanog
42	432	2.9	2665	22	ABG48336	Human liver peptid
43	432	2.9	2665	22	ABP28314	Human peptide #965
44	432	2.9	2665	22	ABP33490	Peptide #996 encod
45	432	2.9	2665	22	ABU18950	Protein #949 encod

ALIGNMENTS

RESULT 1
AAU57452
ID AAU57452 standard; Protein; 2907 AA.

XX AC AAU57452;

XX DT 22-FEB-2000 (first entry)

XX Human transcriptional regulatory factor SEQ ID NO:1.

XX DE Human transcriptional regulatory factor; TCoAl; BLAST detection;
XX KW bromo-domain; cell proliferation; cancer.

XX OS Homo sapiens.

XX PN WO957143-A1.

XX PD 11-NOV-1999.

XX PF 30-APR-1999; 99WO-JP02340.

XX PR 30-APR-1998; 98JP-0137631.

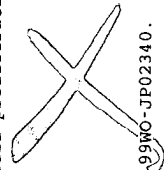
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX PI Jones MH;

XX DR WPI: 2000-052940/04.

XX N-PSDB; AA239032.

XX Transcriptional regulatory factor containing a bromo domain and gene
PT TCoAl encoding it



XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
XX
XX
SQ

Claim 1; Page 48-68; 154pp; Japanese.

The present sequence represents a human transcriptional regulatory factor containing a bromo domain. The factor interacts with proteins involved in the chromatin-mediated transcription regulatory mechanism. It binds to hSNF2H, hSNF2L and NCoA-62/Skip. It can be used for screening compounds binding to it and acting as agonists or antagonists, which are potentially useful for the treatment and prevention of cancer and other cell proliferation disorders.

Query Match 100.0%; Score 14971; DB 21; Length 2907;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYSEEEEDGDAETQSDDEDEDEDEDDSDYPEEMEDDDDDASYCTESSFRSHST 60
DB 1 MYSEEEEDGDAETQSDDEDEDEDEDDSDYPEEMEDDDDDASYCTESSFRSHST 60
QY 61 YSSTGRRKPRVHRSPRIEEDKIDPPLEFPKASSEDLMVNPHEIMNVIAIYEVLRNFGTV 120
DB 61 YSSTGRRKPRVHRSPRIEEDKIDPPLEFPKASSEDLMVNPHEIMNVIAIYEVLRNFGTV 120
QY 121 LRLSPFRFDFCAALVSOQCFLMAEMHVLLKAVLREEDTSNTTGGPADLKDSVNSTLY 180
DB 121 LRLSPFRFDFCAALVSOQCFLMAEMHVLLKAVLREEDTSNTTGGPADLKDSVNSTLY 180
QY 181 FIDGWTWPEVLVAVYCESDKYHHVLPYQAEADYPGPVENKIKVLOFLVDQFLTTNIARE 240
DB 181 FIDGWTWPEVLVAVYCESDKYHHVLPYQAEADYPGPVENKIKVLOFLVDQFLTTNIARE 240
QY 241 ELMSEGVIOYDHCRCVCHKLGDLCCETCSAYVHLECVKPPLEEVPEDEWQCEVCVAHV 300
DB 241 ELMSEGVIOYDHCRCVCHKLGDLCCETCSAYVHLECVKPPLEEVPEDEWQCEVCVAHV 300
QY 301 PGVTCVABIQNKPYIRHEPIGYDRSRKRYWFLNRLLIEEDTENENKKIWIYSTKVQ 360
DB 301 PGVTCVABIQNKPYIRHEPIGYDRSRKRYWFLNRLLIEEDTENENKKIWIYSTKVQ 360
QY 361 LAELIDCLDKYWEAEFLCKILPEMRPEIHRHMDITEDLTNKGNSKSFLEAANEIILES 420
DB 361 LAELIDCLDKYWEAEFLCKILPEMRPEIHRHMDITEDLTNKGNSKSFLEAANEIILES 420
QY 421 IRAKKGIDNVKSPETEEDKNETENDSKAEKNREPEEQSLEKSDSDKTPDDPEQK 480
DB 421 IRAKKGIDNVKSPETEEDKNETENDSKAEKNREPEEQSLEKSDSDKTPDDPEQK 480
QY 481 SEEPTEVGDGKNSVANLGDNTNATSEETSPSEGRSPVGCLETSPDSSNMAEKKVASEL 540
DB 481 SEEPTEVGDGKNSVANLGDNTNATSEETSPSEGRSPVGCLETSPDSSNMAEKKVASEL 540
QY 541 PDQVPEPNKTCESNTSATTTISQPNLENSSSSELNSSQSESAAKADDPNGERESHT 600
DB 541 PDQVPEPNKTCESNTSATTTISQPNLENSSSSELNSSQSESAAKADDPNGERESHT 600
QY 601 PVSIOQIEVGDFTSKSTGSESPGAGKAGSGSTRIITRLRNPDSKLSQKLSQQVAAAA 660
DB 601 PVSIOQIEVGDFTSKSTGSESPGAGKAGSGSTRIITRLRNPDSKLSQKLSQQVAAAA 660
QY 661 HEANKLFGKGEVLVYVNSOGEISRLSTKKEVIMKGNINNYFKLGQEGKRYVHNOYSTNS 720
DB 661 HEANKLFGKGEVLVYVNSOGEISRLSTKKEVIMKGNINNYFKLGQEGKRYVHNOYSTNS 720
QY 721 FALNKHQHRHDHDKRRLAHKFLCTPAGFEKWNKSGVHSGSKVLTISTLRLTITOLENNIPS 780
DB 721 FALNKHQHRHDHDKRRLAHKFLCTPAGFEKWNKSGVHSGSKVLTISTLRLTITOLENNIPS 780
QY 781 SFFHPNASHRANWIKAVQCMCKPREFALALALECAVKPVVMLPIWREFLGHTRLHRMT 840
DB 781 SFFHPNASHRANWIKAVQCMCKPREFALALALECAVKPVVMLPIWREFLGHTRLHRMT 840

QY 841 SIEREEKVKKKKKQEBEETMOQATWVKYTFPVKHQVWKQKGEEYRVTCYGGWSWLSK 900
DB 841 SIEREEKVKKKKKQEBEETMOQATWVKYTFPVKHQVWKQKGEEYRVTCYGGWSWLSK 900
QY 901 THVYRFVPLKPGNTNVNYRKSLIEGTKNMNDENMDESARKCSRPKKIKIEPDESEKDVK 960
DB 901 THVYRFVPLKPGNTNVNYRKSLIEGTKNMNDENMDESARKCSRPKKIKIEPDESEKDVK 960
QY 961 GSDAAKAGADQNMEDISKITERKKDQDVKELLSDSDKPKCEPMEVDDDDMKTESHVNCQES 1020
DB 961 GSDAAKAGADQNMEDISKITERKKDQDVKELLSDSDKPKCEPMEVDDDDMKTESHVNCQES 1020
QY 1021 SQVDVVVYSEGFHLRTSYKKTKSKLDGLLERRIKQFTLEEKQORLEKIKLEGGIKGTGK 1080
DB 1021 SQVDVVVYSEGFHLRTSYKKTKSKLDGLLERRIKQFTLEEKQORLEKIKLEGGIKGTGK 1080
QY 1081 TSTNSSKNLSESPVITTKAKEGCQSDSMROEQSPNANDQPEDLIQCCSQSDSSVLRMSDP 1140
DB 1081 TSTNSSKNLSESPVITTKAKEGCQSDSMROEQSPNANDQPEDLIQCCSQSDSSVLRMSDP 1140
QY 1141 SHTTNKLYPKDRLVDVSIKSPETKCPKONSIENDIEEKVSDLASRGQEPKSKTKGNDF 1200
DB 1141 SHTTNKLYPKDRLVDVSIKSPETKCPKONSIENDIEEKVSDLASRGQEPKSKTKGNDF 1200
QY 1201 FIDDSKLASADDIGTICKNKKPLIOEESDTIVSSSKSALHSSVPKSTNDRDATPLSRAM 1260
DB 1201 FIDDSKLASADDIGTICKNKKPLIOEESDTIVSSSKSALHSSVPKSTNDRDATPLSRAM 1260
QY 1261 DPEGLKGCDSSENSTLENSDTSVSIQDSSEEDMIVQNSNESISEQFRTREODVEVLEPLK 1320
DB 1261 DPEGLKGCDSSENSTLENSDTSVSIQDSSEEDMIVQNSNESISEQFRTREODVEVLEPLK 1320
QY 1321 CELVSGESTGNCEDRLPVKGTANGKPKSOQKKLEERPVPVNCSDQIKLNTTDDKNNENR 1380
DB 1321 CELVSGESTGNCEDRLPVKGTANGKPKSOQKKLEERPVPVNCSDQIKLNTTDDKNNENR 1380
QY 1381 ESEKKGORTSFQINGKDNKPKIYLGECLEKIESESRYVSGNVBPVKNINNKIIPENDIK 1440
DB 1381 ESEKKGORTSFQINGKDNKPKIYLGECLEKIESESRYVSGNVBPVKNINNKIIPENDIK 1440
QY 1441 SLTVKESAIRPFGINDVIMEDFERNESSETKSHLLSSSDAEGNYRDSLETLPSTKESDST 1500
DB 1441 SLTVKESAIRPFGINDVIMEDFERNESSETKSHLLSSSDAEGNYRDSLETLPSTKESDST 1500
QY 1501 QTTTFSACPSNSVNOVEDMEIETSEVKKVTSSPITSEESNLNDRIDENGLPINKNE 1560
DB 1501 QTTTFSACPSNSVNOVEDMEIETSEVKKVTSSPITSEESNLNDRIDENGLPINKNE 1560
QY 1561 NVNGESKRKTIVITEVTTMTSTVATESKTVIKVEKGDQKOTVSVSSTENCASKSTVTTTTVT 1620
DB 1561 NVNGESKRKTIVITEVTTMTSTVATESKTVIKVEKGDQKOTVSVSSTENCASKSTVTTTTVT 1620
QY 1621 KLSTPSTGGSDVIISVKEQSKTVTTVTVDLSLTGGLTVTSMVTSKEYSTRDKVKLMKF 1680
DB 1621 KLSTPSTGGSDVIISVKEQSKTVTTVTVDLSLTGGLTVTSMVTSKEYSTRDKVKLMKF 1680
QY 1681 SRPKKTRSGTALPSYRKFTVKTKKSI FVLPNDDLKLLARKGCIIEVPYFNNAKPALDI 1740
DB 1681 SRPKKTRSGTALPSYRKFTVKTKKSI FVLPNDDLKLLARKGCIIEVPYFNNAKPALDI 1740
QY 1741 WPYPSRPRTFGITWYRLOTVKSLAGVSLMLRLLWASLRWDDMAAKVPPGGSGSTTETSE 1800
DB 1741 WPYPSRPRTFGITWYRLOTVKSLAGVSLMLRLLWASLRWDDMAAKVPPGGSGSTTETSE 1800
QY 1801 TEITITTEIKRRDVGYPYIGIRFEYCIRKICPICGVPEPKETPTPQKGLRSALRPKRPE 1860
DB 1801 TEITITTEIKRRDVGYPYIGIRFEYCIRKICPICGVPEPKETPTPQKGLRSALRPKRPE 1860
QY 1861 TPKQGPVITETWVABEELWEIIRAFARVEKEKAQAVEQQAQKRLKQKFTVATSTT 1920
DB 1861 TPKQGPVITETWVABEELWEIIRAFARVEKEKAQAVEQQAQKRLKQKFTVATSTT 1920
QY 1921 SPTSSTSTSTISPAQKVMVAPISGSVTTGTGKVLVLTTKVSGSPATVTVTQONKNFHQTEATWVK 1980

Db 1921 SPTSSSTTSLSPAQKVMVAPISGVTGKMWLTITKVGSPATVTFQONKNFHOTFTATWK 1980
Qy 1981 QGQNSGVVQVQKVLGIIPSSGTCTSOQFTTSQPTATVTIRPNTSGGGTTSNSQVIT 2040
Db 1981 QGQNSGVVQVQKVLGIIPSSGTCTSOQFTTSQPTATVTIRPNTSGGGTTSNSQVIT 2040
Qy 2041 GPOIRPGMTVIRPPLQOSTLGKAIIRTPVMVQGAQVQVMTQIIRGQPVSTAVSAPNTVS 2100
Db 2041 GPOIRPGMTVIRPPLQOSTLGKAIIRTPVMVQGAQVQVMTQIIRGQPVSTAVSAPNTVS 2100
Qy 2101 STPGQKSLTSATSTSIQSSASOPPRPOQGVKLTMAQLTOLTQHGNGOGLTVVTOGOG 2160
Db 2101 STPGQKSLTSATSTSIQSSASOPPRPOQGVKLTMAQLTOLTQHGNGOGLTVVTOGOG 2160
Qy 2161 QTTGQLQLIPQGVTVLPGGQQLMAAMPNGTVORELFTPLATTATSTTTTTVSTTAA 2220
Db 2161 QTTGQLQLIPQGVTVLPGGQQLMAAMPNGTVORELFTPLATTATSTTTTTVSTTAA 2220
Qy 2221 GTGEQRQKSLSPOMQVHQDKTLPQAQSSVGPAPAKAQPTAQPSARPQPTQPSQAQPEV 2280
Db 2221 GTGEQRQKSLSPOMQVHQDKTLPQAQSSVGPAPAKAQPTAQPSARPQPTQPSQAQPEV 2280
Qy 2281 QTQPEVQTQTTVSSHVPSEAQPTHAQSSKPKVAQSQPOSNVQOGSPVRVQSPQTRIRP 2340
Db 2281 QTQPEVQTQTTVSSHVPSEAQPTHAQSSKPKVAQSQPOSNVQOGSPVRVQSPQTRIRP 2340
Qy 2341 STPSQLSPGQSQVQVQTTQIPIQPHTSIQIPSGQPOQPOVQVQSTOTLSSGQTLNOV 2400
Db 2341 STPSQLSPGQSQVQVQTTQIPIQPHTSIQIPSGQPOQPOVQVQSTOTLSSGQTLNOV 2400
Qy 2401 SVSSSRPQLIQOQPOQVIAVPLQOQVQVLSIQSQVVAQIQAAQSGVPPQIKQLPI 2460
Db 2401 SVSSSRPQLIQOQPOQVIAVPLQOQVQVLSIQSQVVAQIQAAQSGVPPQIKQLPI 2460
Qy 2461 QIQSSAVQTHQIQNVVTVQAASVQEQLOVQOQKQKQOQKQOQKQOQKQOQKQOQKQOQ 2520
Db 2461 QIQSSAVQTHQIQNVVTVQAASVQEQLOVQOQKQKQOQKQOQKQOQKQOQKQOQKQOQ 2520
Qy 2521 EIIQKQVVKHNAVIEHLKQKSMTPAEREENORMIVCQVMKYILDKIDKEKQAKKR 2580
Db 2521 EIIQKQVVKHNAVIEHLKQKSMTPAEREENORMIVCQVMKYILDKIDKEKQAKKR 2580
Qy 2581 KREESVEQRKSONATKLSALLFKHKEQLRAELIKKRALLDKLDQIEVQEEELKRDLIK 2640
Db 2581 KREESVEQRKSONATKLSALLFKHKEQLRAELIKKRALLDKLDQIEVQEEELKRDLIK 2640
Qy 2641 EKDLMLAQATAVAACPVPVLPAPAPPSPPPPPGVOHTGLLSTPTLPVASQKRKR 2700
Db 2641 EKDLMLAQATAVAACPVPVLPAPAPPSPPPPPGVOHTGLLSTPTLPVASQKRKR 2700
Qy 2701 EEKDSKSSKKKMIKSTTQKTDKLYCICKTPYDESKFYIGCDRCQNWYHGRCVGI 2760
Db 2701 EEKDSKSSKKKMIKSTTQKTDKLYCICKTPYDESKFYIGCDRCQNWYHGRCVGI 2760
Qy 2761 LQSEAEILDEYVCPQCOSTEDAMTVLPTTEKDYGLKRVLSLQAHKMAWPLEVPDPN 2820
Db 2761 LQSEAEILDEYVCPQCOSTEDAMTVLPTTEKDYGLKRVLSLQAHKMAWPLEVPDPN 2820
Qy 2821 DAPDYGVVKEPMDLATMEERVQRYEYKLTETVADMTKIFDNCRYNPNSDSPFYQCAEV 2880
Db 2821 DAPDYGVVKEPMDLATMEERVQRYEYKLTETVADMTKIFDNCRYNPNSDSPFYQCAEV 2880
Qy 2881 LESFFVQKLGKFKASRSHNNKLOSTAS 2907
Db 2881 LESFFVQKLGKFKASRSHNNKLOSTAS 2907

RESULT 2
AA57453
ID AA57453 standard; Protein; 2781 AA.
XX
AC AA57453;

XX 22-FEB-2000 (first entry)
XX Human transcriptional regulatory factor SEQ ID NO:10.
DE Human transcriptional regulatory factor; TCoal; BLAST detection;
XX Human; transcriptional regulatory factor; TCoal; BLAST detection;
KW bromo-domain; cell proliferation; cancer.
XX Homo sapiens.
XX WO957143-A1.
XX 11-NOV-1999.
XX 30-APR-1999; 99WO-JP02340.
XX 30-APR-1998; 98JP-0137631.
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX Jones MH;
XX WPI; 2000-052940/04.
XX N-PSDB; AA239033.
XX Transcriptional regulatory factor containing a bromo domain and gene
PT TCoal encoding it
XX Claim 1; Page 139-151; 154pp; Japanese.
XX The present sequence represents a human transcriptional regulatory factor
CC containing a bromo domain. The factor interacts with proteins involved
CC in the chromatin-mediated transcription regulatory mechanism. It binds
CC to hSNF2H, hSNF2L and NCOA-62/Skip. It can be used for screening
CC compounds binding to it and acting as agonists or antagonists, which
CC are potentially useful for the treatment and prevention of cancer and
CC other cell proliferation disorders.
XX Sequence 2781 AA;

Query Match 94.6%; Score 14170; DB 21; Length 2781;
Best Local Similarity 95.1%; Pred. No. 0;
Matches 2766; Conservative 2; Mismatches 13; Indels 126; Gaps 1;
Qy 1 MVSEEEEEEDGDAEETQDSEDEDEEEDDDSDYPEEMEDDDDDAASYCTESSFSRSHST 60
Db 1 MVSEEEEEEDGDAEETQDSEDEDEEEDDDSDYPEEMEDDDDDAASYCTESSFSRSHST 60
Qy 61 YSSTPGRKPRVHRPRSPILEEKDIPPEPKSSEDLMPNEHIMNVIAIYEVLRNFGTV 120
Db 61 YSSTPGRKPRVHRPRSPILEEKDIPPEPKSSEDLMPNEHIMNVIAIYEVLRNFGTV 120
Qy 121 LRUSPREFDFCAALYSQEOCTLMAEMHVLLKAVLREEDTSNTTGPADLKDSVNSTLY 180
Db 121 LRUSPREFDFCAALYSQEOCTLMAEMHVLLKAVLREEDTSNTTGPADLKDSVNSTLY 180
Qy 181 FIDGTMWPEVLRVYCSDKEYHHVLPYQEAEDYPYGPVENKIKVLQFLVDQFLTNIARE 240
Db 181 FIDGTMWPEVLRVYCSDKEYHHVLPYQEAEDYPYGPVENKIKVLQFLVDQFLTNIARE 240
Qy 241 ELMSEGVQYDDHCRVCHKLGDLCCETCSAVYHLECVKPPLEVPDEVOQECVCAHV 300
Db 241 ELMSEGVQYDDHCRVCHKLGDLCCETCSAVYHLECVKPPLEVPDEVOQECVCAHV 300
Qy 301 PGVTDCAVEIQKNKPYIRHEPIGYDRSRKRYWFLNRLLIEEDTENENKKIWYSTKVQ 360
Db 301 PGVTDCAVEIQKNKPYIRHEPIGYDRSRKRYWFLNRLLIEEDTENENKKIWYSTKVQ 360
Qy 361 LAELIDCLDKDYWEALCKILEEMREIHRHMDITDLTNKAGSNKSFLLAANEETLES 420
Db 361 LAELIDCLDKDYWEALCKILEEMREIHRHMDITDLTNKAGSNKSFLLAANEETLES 420
Qy 421 IRAKKGIDNVKSPTEETKDKNETENDSKDAEKNREEFQDSLEKOSDDKTPODDPEQK 480

|||||
Db 421 IRANKGIDNYSPEETEKDNENENDSKAENREFEQOSLEKSDDKTPDDDPQCGK 480
QY 481 SEEPTEVGDGNSVANGLDNTTINATSEETSPSEGRSPVGLSETPDSSNMAEKKVASSEL 540
Db 481 SE-----482
QY 541 PQDVPEPNKTCESNTSATTTSIQPNLENSSSSELNSSQSESAKAADDPENGERSHT 600
Db 483 -----482
QY 601 PVSQTEETVGDFTSEKSTGELSESPGAGKAGSGSTRITRLRNPDSKLSQKSOQVAAAA 660
Db 483 -----VGDFKSEKSNGLSESPGAGKAGSGSTRITRLRNPDSKLSQKSOQVAAAA 534
QY 661 HEANKLFKEGKEVLVNSQGISRLSTKKEVIMKGNINNYFKLGOEGKYRYVHNOYSTNS 720
Db 535 HEANKLFKEGKEVLVNSQGISRLSTKKEVIMKGNINNYFKLGOEGKYRYVHNOYSTNS 594
QY 721 FALNKHQHREDHDKRRHLAHRKFCUPTPAGEFKWNGSVHGSKVLTITSLRLTTIQLENNIPS 780
Db 595 FALNKHQHREDHDKRRHLAHRKFCUPTPAGEFKWNGSVHGSKVLTITSLRLTTIQLENNIPS 654
QY 781 SFFHPNASHRANWIKAYOMCSKPREPALALAILCAVKPVVMLPINREFLGHTRLHRMT 840
Db 655 SFLHPNASHRANWIKAYOMCSKPREPALALAILCAVKPVVMLPINREFLGHTRLHRMT 714
QY 841 STEREEKVKKKKKQEBEETMQOATWVKYTFPVKHQVWKQKGEYRVTCYGGWSWISK 900
Db 715 STEREEKVKKKKKQEBEETMQOATWVKYTFPVKHQVWKQKGEYRVTCYGGWSWISK 774
QY 901 THVYRFVPLPGNTNVNRYKSLGKTKNNMDNDESKRKRCSRPKKTIKIEPDSKEDEVK 960
Db 775 THVYRFVPLPGNTNVNRYKSLGKTKNNMDNDESKRKRCSRPKKTIKIEPDSKEDEVK 834
QY 961 GSDAAKAGDONMDLSKITEKKDOOVKELLDSDSKCKEPEPMEVDDDMKTESHVNCQES 1020
Db 835 GSDAAKAGDONMDLSKITEKKDOOVKELLDSDSKCKEPEPMEVDDDMKTESHVNCQES 894
QY 1021 SOQDVVNVSEGFHLRTSYKKTKSKLDGLLERRIKQFTLEBKQRLKIEGGIKGIGK 1080
Db 895 SOQDVVNVSEGFHLRTSYKKTKSKLDGLLERRIKQFTLEBKQRLKIEGGIKGIGK 954
QY 1081 TSTNSKNLSPEVITTKAKGQCSDSMRQEQSPNANDQPEDLIQGCQSQSDSVLRMSDP 1140
Db 955 TSTNSKNLSPEVITTKAKGQCSDSMRQEQSPNANDQPEDLIQGCQSQSDSVLRMSDP 1014
QY 1141 SHTTNKLYPKORVLDVSTRSPETKCPKQNSITENDIEBKVSDLASRGOEPTKSKTKGDNF 1200
Db 1015 SHTTNKLYPKORVLDVSTRSPETKCPKQNSITENDIEBKVSDLASRGOEPTKSKTKGDNF 1074
QY 1201 FIDDSKLASADDIGTLCKNKKPLTQESDITVSSSKSALHSSVPKSTNDRDATPLSRAM 1260
Db 1075 FIDDSKLASADDIGTLCKNKKPLTQESDITVSSSKSALHSSVPKSTNDRDATPLSRAM 1134
QY 1261 DFEGLKGCDSSENSTLNSDTSVSIQDSSEEDMIVQNSNESISEQOFRTREQDVEVLEPLK 1320
Db 1135 DFEGLKGCDSSENSTLNSDTSVSIQDSSEEDMIVQNSNESISEQOFRTREQDVEVLEPLK 1194
QY 1321 CELVSGESTGNCEDRLPVKGTGANGKPSQOKKLEERPVNKCSQDIKLNKNTTDDKNNENR 1380
Db 1195 CELVSGESTGNCEDRLPVKGTGANGKPSQOKKLEERPVNKCSQDIKLNKNTTDDKNNENR 1254
QY 1381 ESEKKGQRTSTFIQINGKDNKPKIYLGECLEKIESESRYVSGNVPEPKVNNINKIIPENDIK 1440
Db 1255 ESEKKGQRTSTFIQINGKDNKPKIYLGECLEKIESESRYVSGNVPEPKVNNINKIIPENDIK 1314
QY 1441 SLTVKESAIRPPINGDVIMEDEFERNESSETSKHLLSSDADAGNYRDSLETLPSTKESDST 1500
Db 1315 SLTVKESAIRPPINGDVIMEDEFERNESSETSKHLLSSDADAGNYRDSLETLPSTKESDST 1374
QY 1501 QTTTPSASCPESSNVQVEDMEIETSEVKKVTSSPITSEESNLSNDFIDENGLPINKNE 1560
|||||

Db 1375 QTTTPSASCPESSNVQVEDMEIETSEVKKVTSSPITSEESNLSNDFIDENGLPINKNE 1434
QY 1561 NVNGESKRKTVITEVTTMTSTVATESKTVIKVEKGDQOTVVSSTENCAKSTVTTTTTTVT 1620
Db 1435 NVNGESKRKTVITEVTTMTSTVATESKTVIKVEKGDQOTVVSSTENCAKSTVTTTTTTVT 1494
QY 1621 KLSTPSTGGSDVILIISVKEQSKTVVTTTIVTDSLTITGGTLTYSMTVSKEYSTRDKVKLMKF 1680
Db 1495 KLSTPSTGGSDVILIISVKEQSKTVVTTTIVTDSLTITGGTLTYSMTVSKEYSTRDKVKLMKF 1554
QY 1681 SRPKKTSRGTAIPSYRYKPVTKSTKKSIFVLPNDLKKLARKGGIREVPYFYNNAKPALDI 1740
Db 1555 SRPKKTSRGTAIPSYRYKPVTKSTKKSIFVLPNDLKKLARKGGIREVPYFYNNAKPALDI 1614
QY 1741 WPYPSRPRTFGITWRYRLQTVKSLAGVSLMLRLWLASLWMDMAAKYPPPGGSGSTRTETSE 1800
Db 1615 WPYPSRPRTFGITWRYRLQTVKSLAGVSLMLRLWLASLWMDMAAKYPPPGGSGSTRTETSE 1674
QY 1801 TEITTTTEIIKRDRDVPYIGIRFEYCIRKIICPIGYPETPKETPTPQKGLRSSALSURPKPE 1860
Db 1675 TEITTTTEIIKRDRDVPYIGIRFEYCIRKIICPIGYPETPKETPTPQKGLRSSALSURPKPE 1734
QY 1861 TPKOTGPVIEIETWVAEELELWEIRAFARVEKEKAQAVEQOAKKRLQEQKPTVIATSTT 1920
Db 1735 TPKOTGPVIEIETWVAEELELWEIRAFARVEKEKAQAVEQOAKKRLQEQKPTVIATSTT 1794
QY 1921 SPTSTSTSTISPAQKVMVAPISGSVTTGTCKMVLTKVVGSPATVTFQONKNPHQTFATVWK 1980
Db 1795 SPTSTSTSTISPAQKVMVAPISGSVTTGTCKMVLTKVVGSPATVTFQONKNPHQTFATVWK 1854
QY 1981 QGQNSGVVQVOOKVLGIIPSSGTSTSOOTFTSFOPRTATVTIRNPTSGSGGTTNSQVIT 2040
Db 1855 QGQNSGVVQVOOKVLGIIPSSGTSTSOOTFTSFOPRTATVTIRNPTSGSGGTTNSQVIT 1914
QY 2041 GPQIRPGMTVIRTPLOQSTLKGAIIRTPVMVQPAQOQVMTQIIRGQPVSTAVSAPNTVS 2100
Db 1915 GPQIRPGMTVIRTPLOQSTLKGAIIRTPVMVQPAQOQVMTQIIRGQPVSTAVSAPNTVS 1974
QY 2101 STPGOKSLTSTSTNSIOSSASQPPRPOQGOVKLTMAQLTOLTTOGHGNGQLTVVIOQGG 2160
Db 1975 STPGOKSLTSTSTNSIOSSASQPPRPOQGOVKLTMAQLTOLTTOGHGNGQLTVVIOQGG 2034
QY 2161 QTTGOLQILIPQGVTVLPGPGQOLMQAAMPNGTVORFLTPTLATTATTAFTTTTTVTSTAA 2220
Db 2035 QTTGOLQILIPQGVTVLPGPGQOLMQAAMPNGTVORFLTPTLATTATTAFTTTTTVTSTAA 2094
QY 2221 GTGQRQSKLSPQMQVHQDKTLPPAQSSSVGPAKAQPOTAQPSARPQOTQPSAPQPEV 2280
Db 2095 GTGQRQSKLSPQMQVHQDKTLPPAQSSSVGPAKAQPOTAQPSARPQOTQPSAPQPEV 2154
QY 2281 QTPQEVQOTTTVSSHVPSEAOPTHQAQSKKQVAAQSQSQSNVQSQSPRVQSPSQTRIRP 2340
Db 2155 QTPQEVQOTTTVSSHVPSEAOPTHQAQSKKQVAAQSQSQSNVQSQSPRVQSPSQTRIRP 2214
QY 2341 STPSQLSPQSQSOVQTTTSQPIQPIHTSLOIPSGOQPOQPOVOSQSTQTLSSGOTLNOV 2400
Db 2215 STPSQLSPQSQSOVQTTTSQPIQPIHTSLOIPSGOQPOQPOVOSQSTQTLSSGOTLNOV 2274
QY 2401 SVSFSRPLQILQIQPOQVIAVPOLQOQVQVLSIQSQSVVAAQIOAQSQSGVPQOIKLOLPI 2460
Db 2275 SVSFSRPLQILQIQPOQVIAVPOLQOQVQVLSIQSQSVVAAQIOAQSQSGVPQOIKLOLPI 2334
QY 2461 QIQQSSAVQTHQIQNVTVVQAASVOEQLOQVQOLRDQOQKKQOQIEINWVTPSKLLIKV 2520
Db 2335 QIQQSSAVQTHQIQNVTVVQAASVOEQLOQVQOLRDQOQKKQOQIEIKREHTLOAQSS 2394
QY 2521 ELIQQVVMKNNAVITIEHLKOKKMTPAERENORMIVCNQVMKYILDKIDKEEKOAAKR 2580
Db 2395 ELIQQVVMKNNAVITIEHLKOKKMTPAERENORMIVCNQVMKYILDKIDKEEKOAAKR 2454
QY 2581 KREESVEQKRSQKQATKLSALLFKHKEQLRAEILKRALLDKDLQIEVQEBELKRDCLKTK 2640
Db 2455 KREESVEQKRSQKQATKLSALLFKHKEQLRAEILKRALLDKDLQIEVQEBELKRDCLKTK 2514

QY 781 SFPHNASHRANWIKAVOMCSKPREFALALALECAVKKVVMPLPIWREELGHTRLHRMT 840
DB 718 SFHLNPNASHRANWIKAVOMCSKPREFALALALECAVKKVVMPLPIWRESLGHTRLHRMT 777
QY 841 STIEREKEKVKKKKEEETMQOATWVKYTFPVKHQVWKQGBEYRVTOYGGWSWISK 900
DB 778 STIEREKEKVKKKKEEETMQOATWVKYTFPVKHQVWKQGBEYRVTOYGGWSWISK 837
QY 901 THYRFPVKLPONTNWNVTKSLEGTKNMNMENDESRRKRSRPPKKIKIEPDSKDVK 960
DB 838 THYRFPVKLPONTNWNVTKSLEGTKNMNMENDESRRKRSRPPKKIKIEPDSKDVK 897
QY 961 GSDAAKADQNMWDSKITEKKDDQVKELLDSDSKPCKEPEMVEDDDMKTESHVNCQES 1020
DB 898 GSDAAKADQNMWDSKITEKKDDQVKELLDSDSKPCKEPEMVEDDDMKTESHVNCQES 957
QY 1021 SQVDVNVNVEGPHLRTSYKKTKSSKLDGLLERRIKQFTLEBKQRLKIKLEGGIKGICK 1080
DB 958 SQVDVNVNVEGPHLRTSYKKTKSSKLDGLLERRIKQFTLEBKQRLKIKLEGGIKGICK 1017
QY 1081 TSTNSSKNLSESPVITKAKGCOQSDSMRQEQSPNANNQPEDLIOGCSQSDSSVLRMSDP 1140
DB 1018 TSTNSSKNLSESPVITKAKGCOQSDSMRQEQSPNANNQPEDLIOGCSSESQSDSSVLRMSDP 1077
QY 1141 SHTTNKLYPKDRVLDVSIKRSPTKCPKONSTIENDIEEKVSDLASRGQEPKTSKTKGNDF 1200
DB 1078 SHTTNKLYPKDRVLDVSIKRSPTKCPKONSTIENDIEEKVSDLASRGQEPKTSKTKGNDF 1137
QY 1201 FIDDSKLASADDIGTLCKNKKPLIQEESDTIVSSKSAHSSVPKSTNDRDATPLSRAM 1260
DB 1138 FIDDSKLASADDIGTLCKNKKPLIQEESDTIVSSKSAHSSVPKSTNDRDATPLSRAM 1197
QY 1261 DFEGKLGCCSESNSTLENSSDTSVSIQDSSEEDMIVQNSNESISEQFRTREQDVEVLEPLK 1320
DB 1198 DFEGKLGCCSESNSTLENSSDTSVSIQDSSEEDMIVQNSNESISEQFRTREQDVEVLEPLK 1257
QY 1321 CELVSGESTGNCBDRPLVPKGTANGKKPSQOKKLEBPVYNKCSQIKLKNNTDDKNNENR 1380
DB 1258 CELVSGESTGNCBDRPLVPKGTANGKKPSQOKKLEBPVYNKCSQIKLKNNTDDKNNENR 1317
QY 1381 ESEKKGORTSTFOINGKDNKPKIYLGECLEKISESRVYSGNVEPKVNNINKIIPENDIK 1440
DB 1318 ESEKKGORTSTFOINGKDNKPKIYLGECLEKISESRVYSGNVEPKVNNINKIIPENDIK 1377
QY 1441 SLTVKESAIRPFINGDVIEMEDFERNESSETKSHLSSSDAEGNYRDSLETLPSTKESDST 1500
DB 1378 SLTVKESAIRPFINGDVIEMEDFERNESSETKSHLSSSDAEGNYRDSLETLPSTKESDST 1437
QY 1501 QTTTPSASCPESSVNOQVEDMEIETSEVKVKTSSPITSEESNLNDFIDENGLPINKNE 1560
DB 1438 QTTTPSASCPESSVNOQVEDMEIETSEVKVKTSSPITSEESNLNDFIDENGLPINKNE 1497
QY 1561 NVNGESKRKVTIETVMTSTVATESKTIVIKVEKDKQTVVSSSTENCASKSTVTTTTVT 1620
DB 1498 NVNGESKRKVTIETVMTSTVATESKTIVIKVEKDKQTVVSSSTENCASKSTVTTTTVT 1557
QY 1621 KLSPTPGGSDIISVKEQSKTVVTTVTDLSLTGTTGLTVSMVSEKYSTRDVKLMPF 1680
DB 1558 KLSPTPGGSDIISVKEQSKTVVTTVTDLSLTGTTGLTVSMVSEKYSTRDVKLMPF 1617
QY 1681 SRPKKTRSGALPSYRKFVTKSKTSIFVLPNDLKLARKGGIREVPYFNYNAKPALDI 1740
DB 1618 SRPKKTRSGALPSYRKFVTKSKTSIFVLPNDLKLARKGGIREVPYFNYNAKPALDI 1677
QY 1741 WPYSPRPTFGITWRYRLQTVKSLAGVSLMLRLWLASLRWDDMAAKVPPGGGSTRTETSE 1800
DB 1678 WPYSPRPTFGITWRYRLQTVKSLAGVSLMLRLWLASLRWDDMAAKVPPGGGSTRTETSE 1737
QY 1801 TEITTTTIIKRRDVGPIRFEYCIKRIICPIGVPEPKETPTPQKGLRSSALRPKRPE 1860
DB 1738 TEITTTTIIKRRDVGPIRFEYCIKRIICPIGVPEPKETPTPQKGLRSSALRPKRPE 1797
QY 1861 TPQKTGPVLIETWVAEELELWEIRAFERVEKEKAQAVEQAKKLEQOKPTVIATSTT 1920

DB 1798 TPQKTGPVLIETWVAEELELWEIRAFERVEKEKAQAVEQAKKLEQOKPTVIATSTT 1857
QY 1921 SPTSSTTTSTISPAQKVMVAPISGSVTTGKTWLTTKVGSPTVTFQONKNFHOTFATWVK 1980
DB 1858 SPTSSTTTSTISPAQKVMVAPISGSVTTGKTWLTTKVGSPTVTFQONKNFHOTFATWVK 1917
QY 1981 QGQNSGVVOVQOKVLGIIPSSGTSTSOOTFTSFQPRATATVIRPNTSGSGGTTNSQVIT 2040
DB 1918 QGQNSGVVOVQOKVLGIIPSSGTSTSOOTFTSFQPRATATVIRPNTSGSGGTTNSQVIT 1977
QY 2041 GPQIRPGMTVIRTPLOQSTLKGKAIIRTPVMVQGAPOOVMTQIIRGQPVSTASAPNTVS 2100
DB 1978 GPQIRPGMTVIRTPLOQSTLKGKAIIRTPVMVQGAPOOVMTQIIRGQPVSTASAPNTVS 2037
QY 2101 STPCQKSLTSATSTSNIOSSASQPPRPOQGVQKLTMAQLTQLTQGHGNGOGLTVVIQOGQ 2160
DB 2038 STPCQKSLTSATSTSNIOSSASQPPRPOQGVQKLTMAQLTQLTQGHGNGOGLTVVIQOGQ 2097
QY 2161 QTTGQLOLIPOGVTVLPFGPGQOLMQAAMPNGTVORFLFTPLATTATTASTTTTIVSTTAA 2220
DB 2098 QTTGQLOLIPOGVTVLPFGPGQOLMQAAMPNGTVORFLFTPLATTATTASTTTTIVSTTAA 2157
QY 2221 GTGEORQSKLSPQOVHODKTLPPAQSSSVGPAQAQPTQAPSARPOPTQOPSPAQPEV 2280
DB 2158 GTGEORQSKLSPQOVHODKTLPPAQSSSVGPAQAQPTQAPSARPOPTQOPSPAQPEV 2217
QY 2281 QTQPEVQTQTTVSSHVPSEAOPTHAQSSKPOVAAQSQSNVQSGSPVRVQSPQTRIRP 2340
DB 2218 QTQPEVQTQTTVSSHVPSEAOPTHAQSSKPOVAAQSQSNVQSGSPVRVQSPQTRIRP 2277
QY 2341 STPSQLSPGQSOVQTTTQSIPIQTPHTSLQIPSGOGPOQPOQVQSSQTLLSSQTLNQV 2400
DB 2278 STPSQLSPGQSOVQTTTQSIPIQTPHTSLQIPSGOGPOQPOQVQSSQTLLSSQTLNQV 2319
QY 2401 SVSSPSRPLQLIQOQPOQVIAVPQLOQOVVLSQTSQVVAQIOAQOQSGVPPQIKLOLPI 2460
DB 2320 ----- 2319
QY 2461 QIQOSSAVOTHOIQNVTVQAASVQEQBLQRVQOQLRDQOQKKQOQIENVNTPSKLLIKV 2520
DB 2320 ----- 2319
QY 2521 ELIQOVVKNHNAVIEHLKQKKSMTPABREENORMIVCNQVMKYITLDRIDKEEQAAKKR 2580
DB 2320 -----QVVKNHNAVIEHLKQKKSMTPABREENORMIVCNQVMKYITLDRIDKEEQAAKKR 2374
QY 2581 KREESVEQKRKQONATKLSALLFKHKEQLRAELKRALLDKDLQIEVQEBELKRDLEK 2640
DB 2375 KREESVEQKRKQONATKLSALLFKHKEQLRAELKRALLDKDLQIEVQEBELKRDLEK 2434
QY 2641 EKDLMLQAGATAVAAPCPVTPVLPAPPAPPPPPPGVQHTGLLSTPLTPVASQKRKR 2700
DB 2435 EKDLMLQAGATAVAAPCPVTPVLPAPPAPPPPPPPPAVQHTGLLSTPLTPVASQKRKR 2494
QY 2701 EEEKDSSSKSKKKMISTTSKETTKDKTKLYCICKTPYDESK----- 2741
DB 2495 EEEKDSSSKSKKKMISTTSKETTKDKTKLYCICKTPYDESKFYIGCDLCTNMYHCEVGI 2554
QY 2742 -----FYIGCDRCQNWYHGCVCVILQ 2762
DB 2555 TEKEAKMDVYICNDCKRAQEGSSBELYICRTPYDESOQFYIGCDRCQNWYHGCVCVILQ 2614
QY 2763 SEAEILIDEVYCQCOQSTEDAMTVLPLETKDEGLKRVLRSLQAHKMAWPFLEPDPNDA 2822
DB 2615 SEAEILIDEVYCQCOQSTEDAMTVLPLETKDEGLKRVLRSLQAHKMAWPFLEPDPNDA 2674
QY 2823 PDYGVYIKBPMDLATMEERVQRRYIEKLETFEVADMTKIFDNCRRYNNPSSDPFYQCAEVLE 2882
DB 2675 PDYGVYIKBPMDLATMEERVQRRYIEKLETFEVADMTKIFDNCRRYNNPSSDPFYQCAEVLE 2734
QY 2883 SFFVQKLGFKFASRSNNKQSTAS 2907
DB 2907 SFFVQKLGFKFASRSNNKQSTAS 2907

Db 2735 SFFVQKLGFKASRSHNNKLOSTAS 2759

RESULT 4
AAB94078
ID AAB94078 standard; Protein; 557 AA.
XX
XX
AC AAB94078;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:14273.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001
XX 28-30L=2000; 2000EP-0116126.
XX 29-30L=1099; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 14273; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 557 AA;
XX
XX Query Match 18.0%; Score 2695.5; DB 22; Length 557;
XX Best Local Similarity 79.0%; Pred. No. 1.2e-129;
XX Matches 553; Conservative 3; Mismatches 1; Indels 143; Gaps 1;
XX
XX 1937 MVAPISGSVTTGKMWLTTRKVGSPATVTFQNKSFHQTFATVWKQGSNSGVVQVQOKVL 1996

Db 1 MVAPISGSVTTGKMWLTTRKVGSPATVTFQNKSFHQTFATVWKQGSNSGVVQVQOKVL 60
QY 1997 GIIPSTGTSTQQTFTSFQPTATVTIRPNTSGSGGTTNSQVITGPOIRPGMVIPTPLQ 2056
Db 61 GIIPSTGTSTQQTFTSFQPTATVTIRPNTSGSGGTTNSQVITGPOIRPGMVIPTPLQ 120
QY 2057 QSTLGKAIIRTPVMVQPGAPQVMTQIIRGQPVSTAVSAPNTVSTPGQKSLTSATSTN 2116
Db 121 QSTLGKAIIRTPVMVQPGAPQVMTQIIRGQPVSTAVSAPNTVSTPGQKSLTSATSTN 180
QY 2117 IQSSASQPPRPQGGQVKLMAQLTQTOGHGGNGLTVVIGOGGTTGQLQLIPQGVTVL 2176
Db 181 IQSSASQPPRPQGGQVKLMAQLTQTOGHGGNGLTVVIGOGGTTGQLQLIPQGVTVL 240
QY 2177 PGPGQQLMAAMPNGTVQREFLPTATTASTTTTAAAGTGEQKSLSPQMOV 2236
Db 241 PGPGQQLMAAMPNGTVQREFLPTATTASTTTTAAAGTGEQKSLSPQMOV 300
QY 2237 HDKTLPPAQSSSVGPAKAPQTAQPSARPQTPQSPAPQEVQTPQEVQTPQEVQTPQEVQ 2296
Db 301 HDKTLPPAQSSSVGPAKAPQTAQPSARPQTPQSPAPQEVQTPQEVQTPQEVQTPQEVQ 360
QY 2297 PSEAQPTHAQSSKPPQVAAQSQPSQVQSPVQSPQTRIRPSTPSQLSPQSQSQVQT 2356
Db 361 PSEAQPTHAQSSKPPQVAAQSQPSQVQSPVQSPQTRIRPSTPSQLSPQSQSQVQT 420
QY 2357 TTSQPIPIOPHTSLQIPSGQPOQPOVQSSQSTQTLSSGQTLNQVSSVSPSRPQLQIQPQ 2416
Db 421 TTSQPIPIOPHTSLQIPSGQPOQPOVQSSQSTQTLSSGQTLNQVSSVSPSRPQLQIQPQ 446
QY 2417 PQVIAVPOLQVQVLSQISQSVVAGIQAGSQVPPQIKLQPLQIQSSAVQTHQIQNV 2476
Db 447 ----- 446
QY 2477 VTVOASVQEQV 2536
Db 447 -----QVVMKHNNAVIE 457
QY 2537 HLKOKKSMTPAEREENORMIVCNQVMKYILDKIDKEKQAAKKRKEESVEQKRSKONAT 2596
Db 458 HLKOKKSMTPAEREENORMIVCNQVMKYILDKIDKEKQAAKKRKEESVEQKRSKONAT 517
QY 2597 KLSALLFKHKEQLRAEILKRALDKDLQIEVQEEELKRD 2636
Db 518 KLSALLFKHKEQLRAEILKRALDKDLQIEVQEEELKRD 557
RESULT 5
ABB68735
ID ABB68735 standard; Protein; 976 AA.
XX
XX ABB68735;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 32997.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI: 2001-656860/75.
XX N-PSDB; ABL12838.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 32997; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB57737-AB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 976 AA;
Query Match 9.9%; Score 1488.5; DB 22; Length 976;
Best Local Similarity 31.9%; Pred. No. 1.5e-67;
Matches 353; Conservative 139; Mismatches 295; Indels 319; Gaps 21;
4 EEEEEEDGAEETQDSEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHSYSS 63
99 ESEYHYGSDGDEEDKSDNEDDMLLTPSDDESLEVANESESEFSVC---SFNQGV--- 152
64 TPGRRPRVHRPRSPILER-KDIPPLEFPKSSDLMVNPNEHIMNVIAIYEVLNRFTVLR 122
153 ---GRPPRPSPPEVWVLMQEGRYAALDLPDSEDLEFIANTHRLALSIYEVLRRFRHVR 209
123 LSPRFEDCAALVSQEOCTLMAEMHVLLKAVLREEDTSNTTFGPADLKDSVNSTLYFI 182
210 LSPRFEDCAALACEQSALETHVIMLLKALLREEDAQGTFFGLDQKQDVTNLSVLI 269
183 DGMTWPEVLRYVYCESKEY-----HVLVYQEAEDPYGPVENKIKVLQFLVQFLTNIA 238
270 DSITWPEVLRYVYVESDKTDNRNVFHLSTHTE---YPTGIDNRLEVLQFLSDQFLTSNI 326
239 REELMSEGVYQDDHCRVCHKGLDCCCTCSAVHLECVKPPLEVPDEWQCEVCVAH 298
327 RDVNLQEGPIHYDDHCRVCHRGDLCCCTCPAVHLECVDPMDVPTEDQCGCLCRSH 386
299 KVPQVTCVAEIQKNKPYIRHEPIGYDRSRKRYWFLNRLLIIEEDTENENKKIWIYSTK 358
387 KVSQVDCVLPQEQGVLIHRDSLGVDRHGRKYWFARRIFI-EDQENFT---CWYISTT 442
359 VQIAELIDCDKDYWEALCKILEEMREETHRIMDITEDLTNKARGSNKSFLLAANEEL 418
443 SKLLLSRLDAEELTRLSQITRDERDETERQMKLTITNEHKHTKRSVI-----494
419 ESIRAKKGDIDNVKSPETEKNKNETENDSKAEKNREPEFQDSLEKSDDKTPDDPDEQ 478
495 -----ETEQAKNELLEKEVLDED--EKDGDAKSESQIESG 528
479 GKSEPTGVGDKGNSVSANLGDNTNATSETSPSGRSPVGCSETPDSSNMAEKVAS 538
529 TKQDE-----533
539 ELPQDVPEEPNKTCESSNTSATITTSIQPNLENSSSSELNSQESAKAADDPENGERS 598
534 -----533
599 HTPVSIQIEIVGDTSEKSTGELSPGAGKAGSGSTRIITRLNPDNPKLSQKSOQVAA 658
534 -----ECKMVTR-----QKSNQLT- 547

QY 659 AAHEANKLFKEGKEVLVWNSQGEISRLSTKEVKIMGNINNYFKLGQEGKYRYVYNQYST 718
DB 548 -----NGLTFKLMGMEQGFKNYVYNQYST 570
QY 719 NSPALNKHQHRDHRKRLAHKFCUTPAGEFKWNGSVHGSKVLTITLITLITOLENNI 778
DB 571 NPITALNKPQNEERDKRRHLSHKFSLTASDFKWIIGITMTGTTDNNMITTLRQTLINPESNI 630
QY 779 PSSFFHPNASHRANIKAVQCMCKPREFALAILCAVAKPVVWMLPIWREFLGHTRLHR 838
DB 631 AASFNLINWVNNKKNNAVNNARRSEFAVLLLFQASLKSVVFANVHQLGHTTLQR 690
QY 839 MTSTIEREKEKVKKKEKKQ---BEETMOQATWVYTFPVKHQVWQKQGEETRVTYGGW 895
DB 691 ITSAREERKLEKREKREDEERNRLAFNTIKYTLGLKHQVWQKQGEETRVHGWGW 750
QY 896 SWISKT-----HYRVFPKLPNGTNNYKSLGKTKNNMDENDESDDRKCSRPKKIKI 950
DB 751 LMLSSRRRCVRRARQAQLTHNRVYHYTM---GEENDVNEII-----LYDPTQRF 799
QY 951 EPDSEKDEKVGSDAAKAGAQ--NEMDISKITEKKDQDVKELLSDSDSKPCKEPMVEVDD 1008
DB 800 MQCESNVQDQVCHYLPDQYKNVKVIEDVTE-----831
QY 1009 MKTESHVNCQESSQDVVNVVSEGFHL--RTSYKKTKKSSKLDGLLEBRIKQFTLEEKQRL 1066
DB 832 -KIRGH-----IDVSKALNAPGRTYYSKVARKSRLLDLDLRLKLAEEVQOM-- 877
QY 1067 EKILEGGIGKIGTKTSTSSKNLSES 1092
DB 878 -ASKIPSDMKPRLVSSQNNNTANSKQT 902
RESULT 6
AAB54364
ID AAB54364 standard; Protein; 328 AA.
XX
AC AAB54364;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:816.
XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.
OS Homo sapiens.
XX
XX WO2000035320-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05989.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
PI
XX
XX WPI: 2000-579444/54.
XX N-PSDB; AAC99129.
XX
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX

Db 181 DNCRYNPSDS 191

RESULT 8
AAU04323
ID AAU04323 standard; Protein; 238 AA.
XX
AC AAU04323;
XX
XX 18-JUN-1999 (first entry)
XX
XX Human secreted protein SEQ ID NO:69.
XX
XX Human; secreted protein; cancer; tumour; developmental abnormality;
KW foetal deficiency; blood disorder; immune system disorder; inflammation;
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW digestive disorder; endocrine disorder; infection; AIDS.
XX
OS Homo sapiens.
XX
XX WO9910363-A1.
XX
XX 04-MAR-1999.
XX
XX 27-AUG-1998; 98WO-US17709.
XX
XX 29-AUG-1997; 97US-0056271.
PR 29-AUG-1997; 97US-0056073.
PR 29-AUG-1997; 97US-0056247.
PR 29-AUG-1997; 97US-0056270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Pan P, Kyaw H, Rosen CA, Ruben SM, Wei YF;
PI WPI; 1999-190585/16.
XX
XX New isolated human genes and the secreted polypeptides they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
XX Disclosure; Page 18; 170pp; English.
XX
XX AAX30145 to AAX30173 represent 29 isolated human secreted protein genes.
CC AAY04293 to AAY04321 represent the secreted proteins encoded by the 29
CC human genes. The genes and their corresponding secreted polypeptides are
CC useful for preventing, treating or ameliorating medical conditions,
CC e.g. by protein or gene therapy. Also pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new genes. Specific
CC uses are described for each of the 29 genes, based on which tissues they
CC are most highly expressed in, and include developing products for the
CC diagnosis or treatment of cancer, tumours, developmental abnormalities
CC and foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
CC disorders, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The sequences given in AAX30174 to AAX30182 and AAY04322 to AAY04334 are
CC used in the exemplification of the present invention.
XX
XX Sequence 238 AA;
XX
XX Query Match 6.6%; Score 982.5; DB 20; Length 238;
Best Local Similarity 78.8%; Pred. No. 1.6e-42;
Matches 189; Conservative 7; Mismatches 19; Indels 25; Gaps 3;
XX
XX 2689 PTLPVASQKKRREK-----DSSSKSKKKKMIISTSTSKETKK-----DT 2727
3 PTRP----KTPYDESKFYIGCDLCTNWHGECVGIETKAKKMDVYICNDCKRAQGSSE 58

QY 2728 KLYCICKTPYDESKFYIGCDRCQWYHGRVGLQSEAEILDEYVCPQCOSTEDAMTVLT 2787
Db :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 118
59 ELYCICRTPYDESQFYIGCDRCQWYHGRVGLQSEAEILDEYVCPQCOSTEDAMTVLT 118
QY 2788 PLTEKDYEGKLKRVLSLQAHKMAWPFLEPVDNDADPDYGVVKEPMDLATMEERVORRY 2847
Db :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 178
119 PLTEKDYEGKLKRVLSLQAHKMAWPFLEPVDNDADPDYGVVKEPMDLATMEERVORRY 178
QY 2848 EKLTFFVADMTKIFDNCRYNPSDSPFYOCAEVLESFFVOKLKGFKASRSHNNKLOSTAS 2907
Db :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 238
179 EKLTFFVADMTKIFDNCRYNPSDSPFYOCAEVLESFFVOKLKGFKASRSHNNKLOSTAS 238

RESULT 9
AAU19446
ID AAU19446 standard; Protein; 237 AA.
XX
XX AAU19446;
XX
XX 04-DEC-2001 (first entry)
XX
XX Human diagnostic and therapeutic polypeptide (DITHP) #32.
XX
XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW respiratory disorder.
XX
XX Homo sapiens.
XX
XX WO200162927-A2.
XX
XX 30-AUG-2001.
XX
XX 21-FEB-2001; 2001WO-US06059.
XX
XX 24-FEB-2000; 2000US-0184693.
PR 24-FEB-2000; 2000US-0184697.
PR 24-FEB-2000; 2000US-0184698.
PR 24-FEB-2000; 2000US-0184768.
PR 24-FEB-2000; 2000US-0184769.
PR 24-FEB-2000; 2000US-0184770.
PR 24-FEB-2000; 2000US-0184771.
PR 24-FEB-2000; 2000US-0184772.
PR 24-FEB-2000; 2000US-0184773.
PR 24-FEB-2000; 2000US-0184774.
PR 24-FEB-2000; 2000US-0184776.
PR 24-FEB-2000; 2000US-0184777.
PR 24-FEB-2000; 2000US-0184797.
PR 24-FEB-2000; 2000US-0184813.
PR 24-FEB-2000; 2000US-0184837.
PR 24-FEB-2000; 2000US-0184841.
PR 24-FEB-2000; 2000US-0185213.
PR 24-FEB-2000; 2000US-0185216.
PR 12-MAY-2000; 2000US-0203785.
PR 15-MAY-2000; 2000US-0204226.
PR 16-MAY-2000; 2000US-0204525.
PR 16-MAY-2000; 2000US-0204821.
PR 16-MAY-2000; 2000US-0204908.
PR 16-MAY-2000; 2000US-0205232.
PR 17-MAY-2000; 2000US-0204815.
PR 17-MAY-2000; 2000US-0204863.
PR 17-MAY-2000; 2000US-0205221.
PR 17-MAY-2000; 2000US-0205285.
PR 17-MAY-2000; 2000US-0205286.
PR 17-MAY-2000; 2000US-0205287.
PR 17-MAY-2000; 2000US-0205323.
PR 17-MAY-2000; 2000US-0205324.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI

XX	DT	18-DEC-2001	(first entry)	
XX	DE	Novel human secreted protein #2435.		
XX	KW	Human; vaccination; gene therapy; nutritional supplement;		
XX	KW	stem cell proliferation; haematopoiesis; nerve tissue regeneration;		
XX	KW	immune suppression; immune stimulation; anti-inflammatory; leukaemia.		
XX	OS	Homo sapiens.		
XX	PN	WO200179449-A2.		
XX	PD	25-OCT-2001.		
XX	PF	16-APR-2001; 2001WO-US08656.		
XX	PR	18-APR-2000; 2000US-0552929.		
XX	PR	26-JAN-2001; 2001US-0770160.		
XX	PA	(HYSE-) HYSEQ INC.		
XX	PI	Tang YT, Liu C, Drmanac RT;		
XX	PI	WPI; 2001-611725/70.		
XX	DR	Nucleic acids encoding a range of human polypeptides, useful in genetic		
XX	PT	vaccination, testing and therapy -		
XX	PS	Claim 20; Page 529-530; 765pp; English.		
XX	CC	The invention relates to novel human secreted polypeptides. The		
XX	CC	polypeptides and antibodies to the polypeptides are useful for		
XX	CC	determining the presence of or predisposition to a disease associated		
XX	CC	with altered levels of polypeptide. The polypeptides are also useful for		
XX	CC	identifying agents (agonists and antagonists) that bind to them. Cells		
XX	CC	expressing the proteins are useful for identifying a therapeutic agent		
XX	CC	for use in treatment of a pathology related to aberrant expression or		
XX	CC	physiological interactions of the polypeptide. Vectors comprising		
XX	CC	the nucleic acids encoding the polypeptides and cells genetically		
XX	CC	engineered to express them are also useful for producing the proteins.		
XX	CC	The proteins are useful in genetic vaccination, testing and		
XX	CC	therapy, and can be used as nutritional supplements. They may be used to		
XX	CC	increase stem cell proliferation; to regulate haematopoiesis; and in		
XX	CC	bone, cartilage, tendon and/or nerve tissue growth or regeneration;		
XX	CC	immune suppression and/or stimulation; as anti-inflammatory agents; and		
XX	CC	in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid		
XX	CC	sequences of novel human secreted proteins of the invention.		
XX	SQ	Sequence 246 AA;		
Query Match 5.2%; Score 781; DB 22; Length 246;				
Best Local Similarity 67.8%; Pred. No. 3.3e-32;				
Matches 162; Conservative 10; Mismatches 63; Indels 4; Gaps 2;				
Qy	877	HQVWQKGEYRYVTGGWSWISKTHYRFPKLPNTNVNVRKSLGKTKNNMDENDES	936	
Db	7	HQVWQKGEYRYVTGGWSWISKTHYRFPKLPNTNVNVRKSLGKTKNNMDENDES	66	
Qy	937	DKRCSRPKKIIEPDESEKVGSDAAKADONEMDISKITEKKDQDVKELLDSDSK	996	
Db	67	DKRCSRPKKIIEPDESEKVGSDAAKADONEMDISKITEKKDQDVKELLDSDSK	126	
Qy	997	PCKEEPMEVDMDKTESHVNCOESSQVD--VNVNSEGFHLRTSYKKTKSKLDGLLRR	1054	
Db	127	PCKEERTN--GSRXHNRTVCKLSGFEFSRCGCGXXGFSKDXLQENKIIQTRWTSXKE	184	
Qy	1055	IKQTLLEKORLEKIKLEGKIGKGTSTNSSNLSESPVITKAKGCGOSDSMRQSQSP	1113	
Db	185	NXTVYTGKATATRNQVGGWNGXGYRKDSTNSSNLSESPVITKAKGCGOSDSDETRTEP	243	
RESULT 12				
ABB62887				
XX	ID	ABB62887 standard; Protein; 997 AA.		
XX	AC	ABB62887;		
XX	DT	26-MAR-2002 (first entry)		
XX	DE	Drosophila melanogaster polypeptide SEQ ID NO 15453.		
XX	KW	Drosophila; developmental biology; cell signalling; insecticide;		
XX	KW	pharmaceutical.		
XX	OS	Drosophila melanogaster.		
XX	PN	WO200171042-A2.		
XX	PD	27-SEP-2001.		
XX	PF	23-MAR-2001; 2001WO-US09231.		
XX	PR	23-MAR-2000; 2000US-191637P.		
XX	PR	11-JUL-2000; 2000US-0614150.		
XX	PA	(PEKE) PE CORP NY.		
XX	PI	Venter JC, Adams M, Li PWD, Myers EW;		
XX	PI	WPI; 2001-656860/75.		
XX	DR	N-PSDB; ABL06990.		
XX	DR	New isolated nucleic acid detection reagent for detecting 1000 or more		
XX	PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
XX	PT	interactions -		
XX	PS	Disclosure; SEQ ID NO 15453; 21pp + Sequence Listing; English.		
XX	CC	The invention relates to an isolated nucleic acid detection reagent		
XX	CC	capable of detecting 1000 or more genes from Drosophila. The invention is		
XX	CC	useful in developmental biology and in elucidating cell signalling and		
XX	CC	cell-cell interactions in higher eukaryotes for the development of		
XX	CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
XX	CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA		
XX	CC	sequences (ABL01840-ABL16175) and the encoded proteins		
XX	CC	(ABB57737-ABB72072).		
XX	CC	The sequence data for this patent did not form part of the printed		
XX	CC	specification, but was obtained in electronic format directly from WIPO		
XX	CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX	SQ	Sequence 997 AA;		
Query Match 4.4%; Score 661.5; DB 22; Length 997;				
Best Local Similarity 24.7%; Pred. No. 3e-25;				
Matches 309; Conservative 157; Mismatches 395; Indels 389; Gaps 53;				
Qy	1520	DMEIETSEVKVVTSPITSEESNL-----SNDPID---ENGLPINK	1558	
Db	6	NVEIESTDSIVTGLNSGNAEDVDMPGWRRRNRNQKSKSYIGTKDVLDOFLDKDIPLNK	65	
Qy	1559	NENVNGESKRKTVITEVTMTSTVATESKTVIKVEKGDQVWSTENCACAKSTVTTTTT	1618	
Db	66	-----QNRFPPIIT-----ARPVKREC-----VKYERETTFENGNERV-----	97	
Qy	1619	VTKLSTPSTGGSV-----DIISVKEQSKTVVTTTDTSTLTGGLTVTSMVSKYSTPDK	1674	
Db	98	---YSTSPRGVRLNDAAKLYEQA-----VKTEDK	126	
Qy	1675	VKLMEFSRPPKTRSTALPSYRKFVTKTSKISFVLPNDLLKLRKSGIREVPYFNVA	1734	
Db	127	STITK--KPSYR---YPLTSNFTLHKRSLVLPREFELKLARLGGKSTNGFFHAA	180	
Qy	1735	KPALDIWPPSPRPTFGITWRYRLQTVKSLAGVSLMLRLWASLRWDDMAAKVPPGGST	1794	
Db	181	KNN-TIWQCSRPLRFTCWSYTSNATSLSLALQLRLWLSCLRWDDMIK-PPSTDGK	238	

Db 384 -----IDLPSSTEEKSESTAAEDIVKVTTPPEPGSEEDVPKPSQIPEKIT 434
QY 443 ETE-----NDSKDAENREE-----FEDQ-----SLEKSDDKTPDDPEQG 479
Db 435 EDELLKVTSAPAKASPEEVVVTATTSAPTEEDYKPTTAGTISEEBEGKPTP---AEEG 491
QY 480 KSEE-----PTEVGDKNSVANSANLDMTWTNATSEETSPSEGRS-----P 518
Db 492 SGEEEDKVVTAAPEETEDEAKPTSAFY---ASDEKEQEPKPSGSGDDELDLKTPTAPT 548
QY 519 VCCLSETPSSNNAEKKVASELPQDVPB-EPNKTCESSENTSATTTISQPNLENSNSSEL 577
Db 549 AGATGASESEBQDGGK-STEAPTSVDDIEPAKPTSESEASG-----EGEDVAKET 599
QY 578 NSQSSESAAADDPENGERSHTPVSI-OEEIV-GDFTSEKSTGELSSPAGKASGST 635
Db 600 TPAGEASIAEGEEIVKGTTPAGEPSSEGEDEIVKGTTPAESSESEDE----- 648
QY 636 RIITRLRPDSKLSQLKSOQVAAAHAANKLFKECKE--VLVNSQGEISRLSTKKEVIM 693
Db 649 --LTKVTTTTPAGEPSVAGEEIAKETTPAGEIAGEEIVKVTTPAGE--SSIAGEEIVK 705
QY 694 KGNINNYFKLQEGKRVYVHNOYSTNSFALNKHQHREDHDKRHLAHLKFCLTTPAGEKFN 753
Db 706 -----VTPAGESSESSEGEIHK-----VTPAGESSE 734
QY 754 GSVHGSKVLT-----ISTRLITOLENNIPSSFFHPNASHRANKIKA--VOMCSKPRE 806
Db 735 GDEEIVKSTPAGEIAGEEEDVIKATTSAPKSDIEGVKEPETATEVPAEEVEDFAKP-- 792
QY 807 FALALALIECAVPVVMPLTWEEFLGHTRLHRMTSTIEREEKKVKKKKEKKEEETMOQA 866
Db 793 ---TTPIAEEPIAGTPIPTDGI-----GEEIIVKGTTPQPLEQEPSEES 839
QY 867 TWKYTFPPVQHVQWKGGEYRVGYGGWSWISKTHVYFVKLPKGNVNW-YRKSLEGT 925
Db 840 TEV-----PVAED-----DLSSTSSASAIASSTEGV 865
QY 926 KNNMDENMDSDRKRKSRPKIKIBPDSEK-----EVKGSAAKAGADON 971
Db 866 ODAASETSSAPARAGDKDEAATVTAQDKDDEVEQDATDLVEDVQVSTAKTTTTE 925
QY 972 EMDISKITEKQDVKELSDSDKCKEPMVEVDDMK----- 1010
Db 926 QPKSESTEADAEIETVSSPADKOEVPPEAPADKDHKDEEDVQVATDLPKISDGIQPV 985
QY 1011 --TESHVNCQESSQDVVNVYSEGHF-RTSYKKTKSKLGLLERRIKQFTL-----E 1061
Db 986 VDEATTGQPETSETATDKPPSVYLLPPVQSEVPSTAKVNDNRNDETETKPLTPPSGEDQ 1045
QY 1062 EKQLEKIKLEGKIG-----IGKTSNKKSLSESPVITKAKGQCSQSMRQESQNA 1115
Db 1046 SSEPLPAMDLPAGIPGEGDCLVEGKTYANNITVPATAPCDVSK--CISLVACQOQMECK 1103
QY 1116 NNDQPE-----DLIQCCSOS---DSSVLRMSDPSTHTNKLKPDVRLDVSIRSPEKTC 1166
Db 1104 LPENLEKCTVAADLLDGCCTPTICDESTESAEDKDEESTAK--PDNKIDEDVSEISTE-EI 1160
QY 1167 PKQNSIENDIEEK-----VSDLASRGOEPTKSKTKGNDFIDDSKLASADDI 1213
Db 1161 PKVIMPTGITEQPLSHVKPDEEIQPVTVPAPQFDESTAKVKKP--IDES----- 1210
QY 1214 GTLICKNKKPLIOESDITVSSSKSALHSSVPKSTNDRDATPLSRAMDFGKLGCDSESN 1273
Db 1211 ---AEDKKPIGESEED-----SKPIDSEEDKKPVE----- 1238
QY 1274 STLENSDTSVIQSSSEDMIV-----QNSNESISEQFTREOD-----VEVLEPLKE 1322
Db 1239 ---ESAEDKKVPEDESEKPLPTVIPASEIEKESKPEDEKKTADFAATEQDEATTPA 1295
QY 1323 LVSGSTGNCBDRLLPVKGTTEANGK---KPSQOKKLEERPKNKCSQDI---KLKNTTDDKN 1376

1296 QIADTAETAEYDDKLATTAPVSGEDELKPADEKKRTE-----TAQIPDAEIPASTDEPE 1349
1377 NENR---ESEKKGORTST--FQINGKDNKPKIYLKGECLKEISESRVSVGNVPEKVNNI 1430
1350 SSTEELPTVDLUDKPEEDSTKGTAPESDKVPEVPTSASTENEIEEEDSKFTTVPAPKISAS 1409
1431 NKIIP--ENDIKSLTVKE-----SAIRPFINGDVI-----MEDFFERNSSSEKSHL 1474
1410 DETEPTAEEDLVPAATPEIESEFEVSTKPAVOGPPPLPTLAPAQPEKKPVDAETSTEADI 1469
1475 LSSSDAAGNVRDSLETLPSKESDSTQTTTP-SASCPESNSVNOVEDMETEYSEVKKVTS 1533
1470 STEPSAEVEKESGETSESNEIDAGASSTPPVPSADEDKTPTSEKVE---ADDAFTTV 1526
1534 SPIT-SEESN---LSNDFIDENGLPI-----NKNENNVGSK-----RKT 1570
1527 APLAGDEESNLKLPQDIFEEEA-PVAVTTAAAPSKDD---GQKPVVEEKPIEDQGP 1582
1571 VITEVTTMTVATESKTVIKVEKQKQTVVSVSTENCACKSTVTTTITTKLSTPSTGGS 1630
1583 IEDETSTPTS-----SENEIEPESDRATTIAPSKEE--PSEPSTGAPTKDEPAEPSTDA 1635
1631 VDILSVKEQSKTVVTTTVDLSTTGTGLVTSMTVSKYSTRDVKVLMKFRSKPKKTRSGT 1690
1636 ESDES-KETPESEVPTTIVAPA---GEKIPTSSITPDEEPTATSAPVAKPDEDEKETST 1690
1691 ALPSYRKFPVTKSTKKSIFVLPNDDLKLRKGGIREVPYFNNAKPAALDIWYPSPPRTF 1750
1691 EIPTDAPASSEDENS---STDQIPS-----EVP-----EKKP----- 1720
1751 GITWYRLQTVKSLAGVSLMLRLWASLRWDDMAAKVPPGG---GSTRTETSEITEITTE 1807
1721 -----ETAQTPESGDIVGATAAPTSDSEVPVQ 1749
1808 IIKRRDVGPGIRFEYCIRKILICPIGVPEPKETPTPQK-GLRSSALRKRPETPKQTG 1866
1750 RLPE-----EVLAEIPOPSTETGIRKQOODETTAAPSIDAK-- 1783
1867 PVIIETWVAEBELEWEIRAFARVEK--EKAQAVEQ-----OAKKLEQOKPVVIATST 1919
1784 ---EPYVEIDEEAATVAPISEKDEKTEEEKPVEQKPTGEEPESEEEKEKPIEQDVS 1839
1920 TSPTSTSTTISPAQKVMVAFISGVTGTGKMLVLTTKVGSATVTFQONKAFHQTFAWV 1979
1840 EGPVSTEASEAGSTES-----SEEVKPSTEGEVAEKPED----- 1873
1980 KQGQNS-GVVOVQOKVILGIIPSSTG---TSQOFTTSPQRTATVIR-PNTSGSGGTT 2034
1874 KOPSSTAQAPVETIPEISTELPAQDGDKPTSEAPVDSDEDTSAPSDEKIPVSG----- 1927
2035 NSQVITGPIRPGMTVIRTPLOQSTLGRKAIIRTPVMVQFGAPQV-MTQIIRGQPVSTAV 2093
1928 --EEVEGPV-----TTASPOAAEDELKTPAESSEPSSTDKVPETEYQKPEDEKAD 1977
2094 SAPNTVSSPPGKSLTSTATST-----SNIQSASOPPRPOQOQVLTMAQLTQLTQHGCG 2148
1978 ETPESTVQSDVATSTAPVAGDIEKDEQATTASP--EEEEEIKPTIAPAEIPIQ----- 2031
2149 NOGLTVWIOGQTTGQLOLIPQGVTLVLP--GPGQOLMO-AAMPNGTVQVRLFTPLATTA 2205
2032 -----PSEKEPVDE-OEVESGKATPAESDGGQPIDEIATSGPIDE-----ASTA 2076
2206 TTASTTTTIVSTTAAGTGEORQSKLSPQMVHQDK-----TLPPAQSSSVGPAKAQPT 2259
2077 APTKEESTIVASAA-----SP--AVHDEIKDVTITTPQVADEKEVAAPQDETKT 2123
2260 A-----QPSAR-PQOTQOPQSPAQPEVOTQEVQT---QTTVSSHVPS-----EAQPTH- 2304
2124 SIDVSDTSPTAQDDEKQDKTEAPVAPTIVSSSTADSAASSTPEVPSPVEIDKPMDD 2183
2305 --AOSKPKQVA-AQSOPQSNVQOSQPVV--QSPQSTRIRPS-----TPSOLSPQOOSQ 2353
2184 IMSQTIAPHTADGAASTSTDEDEDOAPVTVSPQDAEKTVPSPAPQSDSKTSPSEAPQDADE 2243

2564 RNVKHSFGRRNSTRIASDSQSQPADPLTIKOEHPAQAELKRLQSDSEQKFK 2623
1052 ERRIKQFTLEKORLEKIKLEGGIKIGKTTSTNNKLNSESPVITRAKEQCQSDSMRQEQ 1111
2624 SRHDSNSSIER---KLKTEREI---KTELGDYNSSEYTYTGKLEKESPETRKKHKK 2675
1112 SPNANNDOPEDLOGCSQSDS---VLRMSDPS---HTNKLYPKDRVLDVSI-- 1159
2676 SKRR---LKSSTADTSAAQTPLVMTPLTPSIFDVHSSSECKTFDNEDDLKTEC 2727
1160 ---RSPETKCPKQNSIENDIEBKVSDLASRCQEPKSKTKGNDFFIDDSK 1206
2728 SSIPLEISAGERRHKEKREKLNMWTEATVFN---SPTNDT-----SSE 2773
1207 LASADDITGLCKNKPLIQEESDTIVSSSKSALHSSVPKSTNDRDATPLSR----- 1258
2774 KLSKEERHLKSKSKSMDSNCKTIYNSGSA-H---PSTSPSLPATPTTSAPTAQTSK 2829
1259 ---AMDPEGLGCDSESNTLENSDTSVQDSSEEDMIVONNESISIEQFRTREQDVE 1314
2830 RGEDKMEFTFGIISDEESQPEQATNKDLIPSS---VSTTGPIVSAALQYKQ--- 2881
1315 VLEPLKCELVGESCNGCEDRLPVKGTANGK---KPSOQOKLEERPVNKCSDQIKLN 1370
2882 --EP-----STPNS-----KNEEAHIQLTVHEPEOQOQLER---SRLSGSSSSS 2921
1371 TTDKKNENRESEK---KGQSTFTQINGDKNPKIYKLGCELKEISESRVSGNVEPK 1426
2922 HADREHREKREKREKRSQEQNQIHQKSSK-----VETK 2959
1427 VNNINKIIPENDIKSLTVKESAIRPFINGDVIMEDFERNESSETKSHILLSDDAEGNYR 1486
2960 VDDNSVDMDAERALEAQ-----LMSDFDKPISE----- 2990
1487 SLETLPTSK---ESDSTQTTTPSACPSNSVNOV---EDMEIETSEVKVTSPPIT 1537
2991 --EATESTAATYRSDMTDFRESDN-EDNNSVDMTKQGVKSEQBPQHSKDKKKKKRKS 3047
1538 SEESNLSNDFIDENGLPKNNKENVGESKRTVITEVTMTSTVATESKTVIKVEKGD 1597
3048 EEQKELL-----QOORRESLPNVASTSSAPPTPKGLTVNVOAASK 3088
1598 QT-----VVSSTENCAKSTV-----TTTTTIVKLTSTPGSGVDIIIVSKQS 1640
3089 HADLQDKAHISPPVCKPSPSLPCLIGDDDDALHTPRAKPTTPSSRGNDGLTSPREK 3148
1641 KTV-----VTTVTVDLSLT-----TTGGLTVTS----- 1662
3149 RLISPIPKPTTIANSTLTQSAETPVSSGTVISSSALATTTSTTAAGVSAAPGLDNP 3208
1663 ---MTVSKYST-----RDKVKLMKTSRPPK 1685
3209 TSASAOCKKESFPGDGLDDRESAVQSISAEFNSTSLDNDIADEPKIPVASPPRA 3268
1686 TRSGTALPSYKRVKVTSTKK-----SI 1707
3269 TKPLDKLEESKSRVTISQEEETESAVSALLGESFGTSTTDDYSLDGMDEMSVNELETPTL 3328
1708 FVLPNDDLKKLARK-----GGIREVPVFNNAKPALDIWPPSP-----RPTFGIT 1753
3329 VIAEPDEEALAKAIETAGEPASILEEP-----EMEPERAEAPDPPEAEIESEPVVEVL 3384
1754 WRVRL-QTVKSLAGVSLMLRLLWASLRWDDMAKVPVPGGSGTRETSETTEITTEIKRR 1812
3385 DPEELNAVQSLKHEDMM-----DIKADTPQSERDLQIDT-DTEENPDE---AD 3429
1813 DVGYPGIRFEYKIRKIKICPGVETPKETTPQKGLRSALRPKRPETPKQGPVLIET 1872
3430 SSGP-----SLKIDETVOSSSPEKSSINNS-----PPRETANDI-- 3466
1873 WVAEELELWEIRAFARVEKEKAQAVEQOAKRLEQKDP-----TVIATSTSTPSST 1926

3467 -----PNVESQPKLSNESTPOPSVITKLPFLDTPKTPAGLPSPVKIE 3510
1927 TSTISPAQKVMVAPISGSV---TTGTRKMLVLTTKVGSFATVTFQONKNEHOTFATWVKOG 1982
3511 PPTISKLOPLVQVOTVLPAPHSSTSGISANSVINLDSNVISSCNTSAASATASASA 3570
1983 QNSNGSVVOVKVLGIIPSTSTGTSQGTFTSFQP-RTAIVTIRPNTSGSGGTSN---SQV 2038
3571 SISFGSPTASQNAM---POASTPKOGPIITPOAIRTQSLSIMOPPTISIDEPHPHFAVQOM 3627
2039 ITGPOI---RCQ-MTVIRTLQOQSTL---CKAIIRT-----PVMVQPGAPQ 2078
3628 VLSPOQHHQPPQQTQYMWGIRAFSPHLSHPGVAQSLVGLQSPVGRPMVSP-SPQ 3686
2079 VMTQIIRGQPVSTAVSAPNTVSTPGQKSLTSAATSNIOSSA---SQPPRQOQGVKLT 2135
3687 ---QVQOTQOQHALLITSPQSSNISPLASPTTRVLSSNSPTTSKVNYSQPRNQ----- 3737
2136 MAQLTQLTQCHGNGOGLTVVIQOGQGTGQLOLIPQGVTVLPCPGQOLMOAAMPNGTVQR 2195
3738 -----VPQSPSPKSAEVOQTTPOLMTI---PQKMTPIQVPHIPT-- 3774
2196 FLFTPLATTATTATTTTIVSTTAAGTGEQORSKLSPQMVOHQDXTLPPAQSSSVGPAKA 2255
3775 -IISKVTVQPOQATOSQVASSPPLGS-----LPPHKNVH-----LNAHQN-----QO 3816
2256 QPOT-AQPSARPOPTQPSAPQEPVOTQPEVOTQTVSSHPSEAPTHAOSKSPQVAA 2314
3817 PQVIAKMTAHOHQHMQQFMHQMTQROQHMQ-----QOOLHGQSQ--QITIS 3862
2315 QSQPOSNVGOSPVVQVSPSOTRIRTPSPQLSPGQ-QSOVQTTTSPQIPQIPHTSLQIP 2373
3863 APQOHMQHQAQOQOQHQQHNLQHAQHTQKQHOAQOQFNQ-QIQOQHSQ-- 3918
2374 SQGQPSQPVQS---STQTLSSGQTLNVQSVSPSPQIQIOQPP-QVIAVPOLQOQ- 2428
3919 -QHGVQOQNAQOQHLQSOQHQSOOQLNQ---QHOAQOQOQLOQIQKLOMHGPGQOQKS 3973
2429 -----VOVLSQISQOVVAOIQAOQSGVPOQIKLOLPTQIOQSSAVOTHQIQNV 2476
3974 PQGVGHGSGTSIFASQOHNSQLPAR-----GVPOQ---QHPQOLSHSSPCRPNTLVSV 4024
2477 -VTVOAASVQQLQORVQQLDQOQKQKQ-----QIEINV--NTP 2513
4025 NQGVQPPAILTRVGSHPQNPQOQLPHQSSSGHPHQKOLSSPGANLPLQTPLVNIQNT 4084
2514 SKLLIKVEII-OKV---VMKHNVAIEHLKOKKSWTPAEREENORMIVQNVKYLDKI 2569
4085 -KIIVQQHIVAQNVPPPTQGNATHYPQNGKDSPPGH----- 4123
2570 DKEEQAAKRRKREESVEKRSQONATKLSAL-----LPKHKEQLRAEILKRALLDK 2622
4124 -VEPTPMSAOKTSESVSVIRPTPTTGLAVISANTVGSLLTEENLIKISQPKQDELIEQ 4182
2623 DIQIEVQ-----EELKRDLLKIK-----EKDLMQLAQATAVAAPCPVTPVLPAPAP 2670
4183 DSK-EVDSDYNSAKEVNIQSDSVIKLDTPLASKDAKRAVEMQAI-----AP 4226
2671 PPSPPPPGVQHTGLLSTPLPVASQKREEDSSSKSKKKM 2715
4227 APIPNPQPGNQ--SMAQETALPTTSMVNSNNDHDETEDETETROL 4269

RESULT 15

ABB71160

ID ABB71160 standard; Protein; 5560 AA.

XX

AC ABB71160;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 40272.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 XX N-PSDB; ABL15263.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 PT
 XX Disclosure; SEQ ID NO 40272; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 5560 AA;

Query Match 4.1%; Score 613.5; DB 22; Length 5560;
 Best Local Similarity 19.5%; Pred. No. 9.6e-22;
 Matches 546; Conservative 405; Mismatches 1013; Indels 841; Gaps 118;

QY 305 DCVAEIQKNKPYIRHEPIGVDYRGRKYWFLNR-----RLIIEEDT 344
 DB 1912 DTADKAEKQ---RHEKKEKQEKREKDLRKQVEREEDKDKAKQAEEREKEDKAKEEK 1968
 QY 345 ENENEKKIYYSKVQLAEIDLDYWEAEIACKILEEMREIHRHMDITE-DLTNKA 403
 DB 1969 EREREKKAQSDREK-----KEREEREURE--KEQDKEQKEKEIREKDLREKEQ 2015
 QY 404 GSNKSFLLAAANEIILESIRAKGDIDNVKSPETEEDKDNKNETENDSKDAENREFEEDQSL 463
 DB 2016 -----REDNRREKDLKDLUREKEMREKEQREKELHREKQDREHR 2057
 QY 464 EKSDDKTDPPDDPE--QKSEEPTE-----VGDKGNSVSANLGDNTTNATSEETSPS 513
 DB 2058 EKEQSRAMDVQEGRGMRRELSYQKSKMDIAGEASSITADICQHNKENAMDTIAQGT 2117
 QY 514 EGRSPVGCULSETPDSSNMAEKYVASELPQDVPEPNK---TCESNNTSA-----TTTSI 564
 DB 2118 PGASP-----STP--SDNTPKRSRLSRNSPYRLHKLRLSSQESNHSAGCGGCGSSH 2170
 QY 565 QPNLENSSELNLSNQSSESAAKADDPENGERS--HTPVSIQEEIVGDFTEKSTGELS 622
 DB 2171 QIHEDYVVRIRMENSONSVSHSNQRLNDRNDREKHEKHSKFED-----KNSSSHIS 2223
 QY 623 ESPGAGKAGSGSTRIITRLRNPDSKLSQLKSOQVAAAHAANKLFREKGVLVVNSQGEI 682
 DB 2224 RPHGCGGSSASSSKHHR-----RDKHHQKGSASSIETNSSI-----EYVVDPISQIK 2271

QY 683 SRISTKKEVIMKGINNNYFKLQEGKYRVYHNOYSTNSFALNK-----HOHR----- 729
 DB 2272 HNLNTSEELQSHQ-----PKREKEREHFFSSHANSSRRHKKRHHREKRRHSVA 2324
 QY 730 -----EDHDKRRHLAHKFLCT---PAGEFKWNGSVHGSVKVLTISTLRTITQLENNIP 779
 DB 2325 ESTNTDEEHTPQOHNPHRIRISAAGSGAGSELSSAANTSSGKLHHQHHRSEVERKSSRG 2384
 QY 780 SSFFHPNASHRANWIKAVQMCSPREFALALALECAVPVVMPLPIWREFLGHTRLHRM 839
 DB 2385 DEGHSSKSLRAKLMLSSADSDDTDA-----SKHSI 2419
 QY 840 TSIERE-----EKEKVKKKEKQEEETMOQAATVYKVPVKKQVWVKQGEYRVTCYGG 894
 DB 2420 FDIPTDCPNVSMYDKVKARCKNMQRQAEBKIKAKFS-----QLKSRAKKRKRSTYDG 2474
 QY 895 WS---WISKTHVYRFVPKLPNGTNNVYRKSLCTKNNMDENMDSDKRCRSRPPKIKIE 951
 DB 2475 DSDTEFEDROH-----RNSGSSSFHGR--YPLGSSDDDDDEETHORRIS----- 2517
 QY 952 PDSEKDEVKGSAAKADQNE-----MDISKITEKKDQDVKELLDSDDKPCKEEPMEDV 1006
 DB 2518 -----SPSDAEHGGQDNQAGASTLADANVRQMQ-QNLRRLCDGDDS---SED--EIR 2563
 QY 1007 DMKTESHVNCQESSQVDVVNVSEG-----PHLRTSY-----KKTSKSLDGLL 1051
 DB 2564 RNVMKHSFGKRNSNSTRIASDESQSQAPDLTIKQEHPIAQAQEKREQLSDEQKFK 2623
 QY 1052 ERIKQFTLEKORLEKIKLEGIGKIGTSTNSSKNLSESPVITAKEGCQSDSMRQEQ 1111
 DB 2624 SRHDSNIEER---KLTEREI---KTELGDVYNSSEYTYTGKUYKSPETRRKHKH 2675
 QY 1112 SPANANDQPEDLIQGCQSQSS-----VLRMSDPS---HTTNKLYPKDRVLDVDSI-- 1159
 DB 2676 SKRR-----LKSSSTADTSAQPLVMTPLTPTSFVHSSECKTKEDFNDDLKTEC 2727
 QY 1160 -----RSPETKCPKQNSIENDIEEKVDSLARGQEPKTSKTYGNDFFIDDSK 1206
 DB 2728 SSIPLEISAGERKHKERKEKREKLNRNMTATVPN-----SPTTNDT-----SSE 2773
 QY 1207 LASADDITGLICKNKKPLIQEESDITVSSSKSALHSSVPKSTNDRATPLSR----- 1258
 DB 2774 KLSKEHRHLKSKSKSDNSCNTKIYNSGA-H---PSTPSLPATPTSPASTAQTSK 2829
 QY 1259 ----AMDFEGLKCDSESNSTLNSDTSVSIQDSSEEDMIVQNSNISISQFRTQEDVE 1314
 DB 2830 RGEDKMEFTFGIISDEESQFPEQAETNKDIIPSS-----VSTTGPIVSAALQTYKQ-- 2881
 QY 1315 VLEPLKCELVSGESTGNCEDRLPVKTEANGK-----KPSQOKLEERPVPVKCSDQIKLN 1370
 DB 2882 --EP-----STPNS-----KNEEAHIQTLVHEPEQOQOLER---SRLSGGSSSS 2921
 QY 1371 TTDKKNNENRESEK---KGORTSTFQINGKDNKPIYLKGECLKEISESRVSVGNVEPK 1426
 DB 2922 HADREHRRKREKKREKRSQREQQOQIHOXSK-----LMSDETKPISE-----VETK 2959
 QY 1427 VNNIKIIPENDIKSLTVKESAIRPFINGVIMEDFNERNSSSETKSHLLSSDAEGNYRD 1486
 DB 2960 VDDNSVDMDACRALEAQ-----LMSDETKPISE----- 2990
 QY 1487 SLETLSTK---ESDSTQITTPSASCPESNSVQV-----EDMEIETSEVKKVTSSPIT 1537
 DB 2991 --EATPSTATYRSDMTDVFERSDN-EDNNSVDMTKQGVASEQEQEKSKDKKKRKRK 3047
 QY 1538 SEESNLSNDFIDENGLPIKNENNVGESKRKVITEVTMTSTVATESKTVTKVEKGDK 1597
 DB 3048 EEKQEKLL-----QQORRESLPNVASTSAPPTPGKLTVNVQASK 3088
 QY 1598 QT-----VVSSTENCAKSTV-----TTTTTIVTKLSTPGTGGSDIISYEQS 1640
 DB 3089 HADLQLDAKHISPPVKCPSPSLPCLIGDDDDDALHTPKAKPTTPSSRGNDGLTPSREKP 3148
 QY 1641 KTV-----VTTVTVDSLT-----TTGTGLTVS----- 1662

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 00:43:52 ; Search time 30.6646 Seconds
(without alignments)
9116.791 Million cell updates/sec

Title: US-09-698-295-1

Perfect score:

Sequence: 1 MVSEEEEEEDGDAEETQDSE.....KLKGFKASRSHNKLQSTAS 2907

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 2833308

Minimum DB seq length: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Minimum Match 0%	Maximum Match 100%
0	100
1	100
2	100
3	100
4	100
5	100
6	100
7	100
8	100
9	100
10	100
11	100
12	100
13	100
14	100
15	100
16	100
17	100
18	100
19	100
20	100
21	100
22	100
23	100
24	100
25	100
26	100
27	100
28	100
29	100
30	100
31	100
32	100
33	100
34	100
35	100
36	100
37	100
38	100
39	100
40	100
41	100
42	100
43	100
44	100
45	100
46	100
47	100
48	100
49	100
50	100
51	100
52	100
53	100
54	100
55	100
56	100
57	100
58	100
59	100
60	100
61	100
62	100
63	100
64	100
65	100
66	100
67	100
68	100
69	100
70	100
71	100
72	100
73	100
74	100
75	100
76	100
77	100
78	100
79	100
80	100
81	100
82	100
83	100
84	100
85	100
86	100
87	100
88	100
89	100
90	100
91	100
92	100
93	100
94	100
95	100
96	100
97	100
98	100
99	100
100	100

Listing first 45 summaries

Database : PIR 76:

1: pirl

```
1: pir1
2: pir2
```

3: pir3

4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	4082	27	3	810	2	G01252
2	854.5	5.7	7	1711	2	T21432
3	546.5	3.7	537	2	T13564	small GTP binding microtubule-protein hypothetical protein
4	514.5	3.4	2447	2	T16870	hypothetical protein
5	517.5	3.3	5170	2	T15348	hypothetical protein
6	463.5	3.1	3575	2	T34513	hypothetical protein
7	462.5	3.1	3759	2	A35085	trithorax protein
8	458.5	3.1	3488	2	T34418	hypothetical protein
9	446	3.0	2897	2	B48666	cell proliferation
10	440	2.9	2271	2	F90073	hypothetical protein
11	439.5	2.9	3256	2	A48666	cell proliferation
12	439.5	2.9	3924	2	S37431	ankyrin 2, neuronal
13	439	2.9	2938	2	T30249	cell proliferation
14	419.5	2.8	2232	2	T34434	hypothetical protein
15	417	2.8	2841	2	D71623	erythrocyte membrane
16	407.5	2.7	3828	2	T13857	trithorax protein
17	406	2.7	4377	2	A55575	ankyrin 3, long sp
18	405.5	2.7	452	2	T21435	hypothetical protein
19	403	2.7	5105	2	T32650	hypothetical protein
20	402.5	2.7	510	2	T21430	hypothetical protein
21	401.5	2.7	405	2	T21433	hypothetical protein
22	401.5	2.7	1367	1	S48478	glucan 1,4-alpha-g
23	397	2.7	2722	2	T20532	hypothetical protein
24	397	2.7	6642	2	T29757	protein UNC-89 - C
25	393	2.6	1630	2	A35577	ascites stialoglyco
26	393	2.6	2481	2	D90011	PmtB protein (imp
27	392	2.6	1871	2	D96796	probable heat shock
28	392	2.6	2738	2	E88320	protein F07A11.6 [
29	391	2.6	1939	2	T18372	repeat organellar

RESULT 1

REPORT
G01252

small GTP binding protein SEC4 homolog - human

C;Species: Homo sapiens--(man)

C:\Specres\homo\suprens (man)
C:\Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 29-Jan-1999

C; Accession:—G01252

R; Bowser, R. P.

submitted to the EMBL Data Library, January 1994

A;Reference number: G064411

A;Accession: G01252

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-810 <BOW>

A;Cross-references: EMBL:U05237; NID:g1276427; PID:g451848

C;Genetics:

A;Gene: FAC1

Query Match 27.3%; Score 4082; DB 2; Length 810;

Best Local Similarity $\sim 98.73\%$ Pred. No. 3.2e-144;

Matches	777	Conservative	4	Mismatches	2	Indels	4	Gaps	1
<p> </p>									

Qy	1	MVSEEEBEGDABETQDSEDEDEMEBDDDDSDYPMEBDDDDDDASYCTESSFRSHST	60
Db	1	MVSEEEBEGDABETQDSEDEDEMEBDDDDSDYPMEBDDDDDDASYCTESSFRSHST	60
Qy	61	YSSTPGRKKPRVHRPSRPILEEKDIPLEPPKSSEDLWPNHEIMNVIAIYEVLRNFGTV	120
Db	61	YSSTPGRKKPRVHRPSRPILEEKDIPLEPPKSSEDLWPNHEIMNVIAIYEVLRNFGTV	120
Qy	121	LRLSPPFFEDFCAALVSEQECTLMAEMHVLLKAVLREEDTSNTTFGADLKDSVNSTLY	180
Db	121	LRLSPPFFEDFCAALVSEQECTLMAEMHVLLKAVLREEDTSNTTFGADLKDSVNSTLY	180
Qy	181	FIDGMTWPEVLRYVCESDKHHVLPQOEADIPYGPVENKIKVLQFLVQDFLTNTIARE	240
Db	181	FIDGMTWPEVLRYVCESDKHHVLPQOEADIPYGPVENKIKVLQFLVQDFLTNTIARE	240
Qy	241	ELMSEGVIOYDDHCRVCHKLGDLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKV	300
Db	241	ELMSEGVIOYDDHCRVCHKLGDLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKV	300
Qy	301	PGVTDCAEATQKNKPYTRHEPIGYDRSRRYWFNLRRLLIIEEDTENENEKKIWYYSKQV	360
Db	301	PGVTDCAEATQKNKPYTRHEPIGYDRSRRYWFNLRRLLIIEEDTENENEKKIWYYSKQV	360
Qy	361	LAELIDCLDKDYWEAECLKILEENRESETHRMDITEDLTNKARGSNKSFLAAANEETILES	420
Db	361	LAELIDCLDKDYWEAECLKILEENRESETHRMDITEDLTNKARGSNKSFLAAANEETILES	420
Qy	421	IRAKKGIDINVKSPTEETKDNKTENDNSKDAEKNREEFEDQSLKESDDKTIPDDDPQGGK	480
Db	421	IRAKKGIDINVKSPTEETKDNKTENDNSKDAEKNREEFEDQSLKESDDKTIPDDDPQGGK	480

QY	481	SEPTVEYDGNKSVNSANLGDNTTNAATSEETSPSEGRSPVGCSETPDDSSNMAEKKVASL	540
Db	481	SEPTVEYDGNKSVNSANLGDNTTNAATSEETSPSEGRSPVGCSETPDDSSNMAEKKVASL	540
QY	541	QDVPEPNKTCSSNTSATTTIQPNLENSNSSELSSQSESASAKAADDPENGERSHT	600
Db	541	QDVPEPNKTCSSNTSATTTIQPNLENSNSSELSSQSESASAKAADDPENGERSHT	600
QY	601	PVSTQEEIVGDFTSEKSTGELSPGAGKAGSGASTRIITRLRNPDSKLSQKSOQVAAAA	660
Db	601	PVSTQEEIVGDFTSEKSTGELSPGAGKAGSGASTRIITRLRNPDSKLSQKSOQVAAAA	660
QY	661	HEANKLFKEGKEVLVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKYRVYHNOYSTNS	720
Db	661	HEANKLFKEGKEVLVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKYRVYHNOYSTNS	720
QY	721	FALNKHQHRDHDKRRHLAHLKFCITPAGEFKWNGSVHGSVKVLTISTRLTITQLENNIPS	780
Db	721	FALNKHQHRDHDKRRHLAHLKFCITPAGEFKWNGSVHGSVKVLTISTRLTITQLENNIPS	780
QY	781	SFFHPNW 787	
Db	777	TSLHPSF 783	
RESULT 2			
T21432			
hypothetical protein F26H11.2 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999			
C:Accession: T21432			
R:Barlow, K.			
submitted to the EMBL Data Library, November 1996			
A:Reference number: 219421			
A:Accession: T21432			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-1711 <WIL>			
A:Cross-references: EMBL:Z81515; PIDN: CAB04197.1; GSPDB: GN00020; CESP: F26H11.2			
A:Experimental source: clone F26H11			
C:Genetics:			
A:Gene: CESP:F26H11.2			
A:Map position: 2			
A:Introns: 56/3; 154/2; 373/2; 417/3; 505/2; 596/1; 703/2; 847/3; 922/3; 1034/2; 1078/2;			
Query Match			
Best Local Similarity 17.3%; Score 854.5; DB 2; Length 1711;			
Matches 431; Conservative 296; Mismatches 624; Indels 1139; Gaps 72;			
QY	4	EEEEEE---DGAEEETQDSDE---EDEMEEDDDSDYPPEMEDD---DDDASYCTE	52
Db	169	KROEEDIIYMDSEDESEESSDDEFLMNEQVQVEEELNLTIDIKIEKGLDENKYC--	226
QY	53	SFRRSHSYSTPGRRKRVHRPRSPILKE--KDIPLEPKSSSEDLWVNEHIMVIAI	110
Db	227	-----PWLDEPASPILKLEPSESSQDIPITASIMDAVEI	261
QY	111	YEVLRNFTGLRLSPFRFEDCAALVSOEQCTLMAEMHVLLKAVLREEDTSNTTFPAD	170
Db	262	YELRSYHRTLRIPTFTFEDFCAALISHNNSCIAEVAHMLLRNCLKSDDEQTHYSVTE	321
QY	171	LKDSVNSTLYFDGMTWFEVLVRVYCESDKYHHVLPQEA-----EDYPGPVENKI	222
Db	322	TNNSVINIHMMDTLTYAEILRQYEA-----YPFADASVRDAINVDNTPFVGVDANI	374
QY	223	K-----VLQFLVDOFLTNTIARELSMSESVIQYDDHCRVCHK--LGDLLC	265
Db	375	QRDFSEFFNIKHVFLVLLVFWYSFELYSSEFKLVNNGVGFQDENCRCVGKSGSRGVG	434
QY	266	CETCSAVYHLECVKPLBEVPEDEMQCEVCAHK--VPGVTCVAVIAIOKNKPIYIHEPIGY	324
Db	435	CTQCEAAAFHVEC--SHLKPFPE-VLVCNICKNSAVRGVLPPPEAVDREP--LRSQPIGR	489
QY	325	DRSRKYYWFLNRRLLIIBEDTENENEKKIWIYSTVKVQLAELIDCLDKDYWEAELCKILEEM	384
Db	490	DRYGRYWFIVRRLVWQSLDETE---LYYSTVPQLYQLLQKLDRTYYEKDCLDCTIRLR	545
QY	385	REEIHRHMDITEDITNARGSNKSFALAAANEIILESIRAKK---GDDINVKSPSETEKDK	441
Db	546	IDFLEQALVEMTSERR-----EALLETWKVQLIIGYDFAEATTPQ-----	588
QY	442	NETENDSKDAEKREEFEDQSLKSDDDKTPD---DDPEQKSEETPEVGDGKNGSVSANL	498
Db	589	-----IYLHRDSMKRMASILRDCAQAGOVQKQEVKL-----	618
QY	499	GDNNTNATSEETSPSEGRSPVGCSETPDDSSNMAEKKVASLPODVPEEPNKTCESSNTS	558
Db	619	-----EE--PVGEQSPVKCVQFVEDSI-----LPE-----	641
QY	559	ATTTIQPNLENSNSSELSSQSESASAKAADDPENGERSHTPVSIOEIVGDFTSBKST	618
Db	642	-----SMGIFDAKLINTFWSGGA-----TOEELVEQFV---DI	672
QY	619	GELSESPGAGKAGSGASTRIITRLRNPDSKLSQKSOQVAAAAHEANKLFKEGKEVLVNS	678
Db	673	SDNFDAPSA-----	681
QY	679	QGEISRLSTKKEVIMKGNINNYFKLGQEGK---YEVYHNOYSTNSFALNKHQHRDHDKR	735
Db	682	-----NLWRMGDEGNDQTFMTYINYSRNMDSSEFTTRKKAADKK	721
QY	736	RHLAHLKFCITPAGEFKW---NGSVHGSVKVLTISTRLTITQLENNIPSSFFHPNASHR	791
Db	722	KYMASKF--AQIDNEDVWVAKNRQFYGDASLHCKFIMWTLOQVINKNIPIDLHMRKWPEFA	779
QY	792	ANWIKAVOMCXPREFALALILECAVPVVMPLPWREFLGHTRLHRMTSTEREKEKVK	851
Db	780	KGFDLEVSADDDYKKLVTCLLKLDCAVRKTIIMPQWNGLGOTRLERTIVDQRENFMKEQ	839
QY	852	KKEKK-----QEEETMQQATWYKTYFPVKHVKOKGGEYRVTG---YGGWSWISK	901
Db	840	QLKKIDADALTKDLDSEFVRVYNYKPKWPNTY--ILRQGEYRNAGKSGGAWAAK	898
QY	902	HVYRFVPLPGNTNVNYRKSLGKTKNMNDENDESDESKRCSPPKIKIEDPSEKDEKVG	961
Db	899	YVEKWI-----QVPESPKPLPLAVTVEIKTESVSNRK-----	930
QY	962	SDAAKADQNEMDISKITEKKDDQVKELLDSDSKPCKEPEMEVDMDKMTESHVNCQESS	1021
Db	931	-----ARLELLVSKIT-----	942
QY	1022	QVDVNVVSEGFLRTSYKKKTKSKLDGLLERRIKQFTLEEKQRLKIKLEGGIKIGIKT	1081
Db	943	-----KKQRSG-----GKS	952
QY	1082	STNSSKNLSESPVITKAKGQSDSMRQEQSPNANNQDPEDLIQCSQSDSVLRMSDPS	1141
Db	953	SKKTPFELT-----NGCYSPCRS-----	971
QY	1142	HTTNKLYPRDVRLLDDVSIRESPTCKPKQNSIENDTEKVSDLASRQGEPTKTKGNDFF	1201
Db	972	-----NPNRKC-----YSP-----	980
QY	1202	IDDSKLASADDIGTLICKNKKPLIQEESDTIVSSSKSALHSSVPKSTNDRDATPLSRAMD	1261
Db	981	-----MCRN-----	984
QY	1262	FEGLGCDSESNSTLENSSDTVSIQDSSEEDMIVQNSNESISEQFTRREQDVEVLEPLKC	1321
Db	985	-----	984
QY	1322	ELVSGESTGNCEDRLPVKGTPEANGKKPSQOKKLEERPVNKCSQIKLKNNTDKKNENRE	1381
Db	985	-----GYLVSAKQAHDERKLEE-----	1001

```
Qy 1382 SEKKQRTSTFQINGKONKPKIYKLGECLEKSEISRVVSGNVKPKVNNKIIPENDIKS 1441
Db 1002 -----
Qy 1442 LTVKESAIRPFINGDVIMEDFERNSETSKSHLLSSDAEGNYRDSLETLPSTKESDSTQ 1501
Db 1002 -----
Qy 1502 TTTTSPASCPESNVQVEDMEIETSEYKVKVTSSPITSEESNLNDFIDENGLPIKNEN 1561
Db 1002 -----SG 1003
Qy 1562 VNGESKKKTVITEVTTMTSTVATESKTVIKVEKDKQTVVSSSTENCAKSTVTTTTVTVK 1621
Db 1004 VLGEK-----
Qy 1622 LSTPSTGSDIISVKEQSKTVVTTTDSLTITGGTLVTSMTVSKYSTRDKVKLMKFS 1681
Db 1010 -----
Qy 1682 RPKTRSGTALPSYRKFTVKTSKKSIFVLPNDLKLARKGGIREV--PYFNYNNAKPAID 1739
Db 1010 -----AWPIEIQETSTKRGKSIQVLOKKILROMIMGCGCOQVYMPGFSAGIKSNLL 1062
Qy 1740 IMPYSPRPFTGITWRYLQTVKSLAGVSLMLRLWASLWDDMAAKVPPGGSGSTRTETS 1799
Db 1063 IMPYAPRPTLDLCWKQTLNARSLHALQKLIWSSIKFNEFD---PDDTHPDRRWI 1119
Qy 1800 ET--EITTEIKRQDVPYRFEYCIKIIQPI-GVPTPKETPTPQRKG-----LR 1850
Db 1120 DTSHDERRLIHKEPPYQOYERYEMEIEIPLYDEPEEDESWLSRNGSSSESHR 1179
Qy 1851 SSALRPKRPT-----PKQT-----GPVITWVAEELELWE----- 1883
Db 1180 SSSARKRPORHEFLSLKFGNPKKSNFSLNRRATAIRREWDVGVTLKVFEEKDYW 1239
Qy 1884 --TRAFARVEKEKAQV-----EQAKRLEQOQFTVIATSTTSPTSTSTISPAQ 1934
Db 1240 KWIHAEAKTAKRLEATRAKAKAKADEERRRQQOORSVARIPV-PMHS-----LIPSE 1294
Qy 1935 KVMYAPISGVS---TTGTRKVLTKVGSPTATVFOQN-----KNFHOTFATWVKQG 1983
Db 1295 RNNV-PYLGSOQORRPNNGERGFLEKYNSSSVSPQAHGYASTPPGCHQPPQNIIRQAG 1353
Qy 1984 SNGSVVQVQKVLGIIPSSGTSTGQFTTSQPTATVIRPNTSGSGGTTSSNQVITGPQ 2043
Db 1354 YNQ-----LPRKPTTSPFNFS-RP-VATITPTPOLRAAAGADG---VVRAMV 1396
Qy 2044 IRPG---MTVIRTPLOQSTLGRKAIIRTPVMVQPGAPQVMTQIIRGO----- 2087
Db 1397 MTPGNKSTVNTNTPYPQ-----ALNRQOYQLOQOQOQOQAVRRLTNGYHFMGDMRGGGR 1451
Qy 2088 -----PVSTAVSAPNTVSTPGQKSLTSAT-----STSNTOSS 2120
Db 1452 NPSVQMQRHLQPNRAALQRFEGESTTEMRRVTEAAIPDNDGDEQPPVPIPYDPTSNFDAQ 1511
Qy 2121 ASQPPRQOQGVKLTMAQLTQLTGHGGNGLTVVIOGQGTGQLQL-----IPQGVTVL 2176
Db 1512 RAQOQHQPQRPVSTPAQMIRTTQPGGVKH--NVILMKASDGTQKMWLKPQPPPGTVI- 1568
Qy 2177 PGQOQLMQAAMPNGTVORELFTPLATATATTTTSTTAAGTGEQRQSKL----- 2230
Db 1569 -STGQVVVVRQPTAVQORQLYT-----ATPGTRVVRIPNANGAPRQDDHVMRRVYQ 1621
Qy 2231 --SPQMVQHDK--TLPPAQ-----SSSVGPAKAQPOQAQPSARPQPOQPOS 2274
Db 1622 ASGRAMEYDDOGTPPGQOVRYVLOGGNSGTPNPNPKVSSRGGPRGLTMQWVQOQQ 1681
Qy 2275 PAQREVQTOPEVOT---QTTSSHVPSEAQ 2301
Db 1682 QHNPAHYDMPDDATGFAVSTTTTQVDPDEQ 1711
```

RESULT 3

T13564

microtubule-associated protein homolog - fruit fly (*Drosophila melanogaster*)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: *Drosophila melanogaster*
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.
A:Reference number: Z17689

A:Accession: T13564

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5327 <SPA>

A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1

C:Genetics:

A:Cross-references: FlyBase:FBgn0025392

A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1

A:Note: EG:49E4.1

C:Superfamily: *Drosophila* 576K microtubule-associated protein homolog

Query Match

3.7%; Score 546.5; DB 2; Length 5327;

Best Local Similarity 18.3%; Pred No. 3.9e-12;

Matches 550; Conservative 479; Mismatches 1134; Indels 825; Gaps 121;

Qy 2 VSEEEEDGDAAETQDS-----EDDEE-----DEMEEDDDSDYPPEMEDDD 44

Db 2191 VKDEAQESRRSKTESVPPKAKDDKSPKEVLQPVYSMTETREDADQPMKPSQASRR 2250

Qy 45 DD-ASYTESFSRH-----STYSSTPGR-----RKPRV-----HRPSPILEE 82

Db 2251 ESTAESIKASPRDEKSPKASKEASRFGSVAESIKYDLDPKQIIKDDKSTHSRESLED 2310

Qy 83 KDIPPLPEPKS-SEDLMPVNEHIMNVIAIYVLNFGVTLRLSPERED-----FCAALV 136

Db 2311 KSAVTS--KSVRPLSVASDH-PAVAIEDDAKS-----SISP---KDKSRPGFVAETV 2359

Qy 137 SQ--EQCTLMAEMHVLLKAVLR-----EEDTSNTFGPADLKDSVNSTLYFID 183

Db 2360 SSPIEATMEFSEKTEVVEKSSALSLQGGSGKQTDSSPVDVAEGDFSHVASVSTVTP 2419

Qy 184 GMTWP-EVLRVYCESDKEYHHVLPYQAEADYPYGP-----VENKIKVLQ 226

Db 2420 TLTKPAELAQIGAAKTVS---SPLDEALRTPSAPEHISRADSPAECASEIASQDKSPQ 2475

Qy 227 FLVDOFTTNIARELSSEGVIQYDDHCRVCKLGLDLCCTCSAVYHLECVK- 280

Db 2476 VLKSSRPANVAESK-----DDAQLKSSVEDLR-----SPVASTEISRASAGET 2521

Qy 281 ---PLEEVPDEMOCEYVAHKVPGVTDCAVEIQKNKPYIRHEPIGYDRSRKKYWFNLNR 337

Db 2522 ASSPIEAPKDFAEFE--QAEK--AVLPLTIELKGNLP----- 2555

Qy 338 LIIEEDTENENEKKIWTYSYTKVQLAELIDCLDKDYWEALCKLILEEMREEIHRHWDITED 397

Db 2556 -----TLSSPVDVAHASVQPAELSKV-----DIKT 2581

Qy 398 LTKARGSNKSLAAANEETILESIRAKKGDIDNVKSPETEKKDKNETENDSKDAEKNEE 457

Db 2582 ASSPIDCAPKSLIGSPAE-----RPESPAESAKDAESVEKSKDASRRPSV 2628

Qy 458 FEQSLKSDSD-----KTPDDDPQEQK-SEETPEVGDKGNSVANL-GDNTNAT 506

Db 2629 VESTKADSTKGDISPSPESVLEGGPKDDVEKSKSSRPP-----SVSASITGDKDVS 2681

Qy 507 -----SEETSPSEGRSPVCLSETPDSNNMAEKKVASLQDVPPEPNKTCSSN 556

Db 2682 RPASVWESVKDEHDKAESRRRESIAKVESVIDEAGSKSKSS---QDSQDKSTLASKE 2738

Qy 557 TSATTTTIQPNLENSSSSELSNQSSESAKAADDPENGERSHTPVSIQE-----EIVGD 611

Db 2739 ASRESVY----ESSKDDAKESRPESVIASGEPV--PRESKSPLDKSTSRPGSVVES 2792

Db 4579 EEPADVDERSVKESKSIATIMTTSIIPKSEDM--EPISK--LVEEBHE-HVEELAQ 4633
QY 2634 RDLIKIKKDLMLAQATAVAAPCPVPVYL-----PAPPAPPPS 2673
Db 4634 EVTSTKTTLLQSSQSSTTSSTKTGASRVESITLTQMDQQTQSOGDPADKTPPT 4693
QY 2674 PPPPGVQ 2681
Db 4694 APVSPGVK 4701

RESULT 4
T16870
hypothetical protein T13H2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Sep-2000
C:Accession: T16870
R:Wu, X.
submitted to the EMBL Data Library, October 1995
A:Description: The sequence of C. elegans cosmid T13H2.
A:Reference number: Z18593
A:Accession: T16870
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2447 <NUX>
A:Cross-references: EMBL:U39653; NID:g1049397; PID:g1049400; PIDN:AAB52494.1; GSPDB:GN00
A:Experimental source: strain Bristol N2; clone T13H2
C:Genetics:
A:Gene: CESP:T13H2.3
A:Map position: X
A:Introns: 18/3; 135/3; 176/3; 496/1; 693/1; 816/1; 1053/1; 1131/3; 1181/3; 1308/3; 1645
C:Superfamily: RING finger homology
F:158-207/Domain: RING finger homology <RRN>

Query Match 3.4%; Score 514.5; DB 2; Length 2447;
Best Local Similarity 18.2%; Pred. No. 2.2e-11;
Matches 514; Conservative 423; Mismatches 1037; Indels 857; Gaps 118;

QY 9 EDGAEETQDSEDEMEDEDDSDYPPEMEDDDDDASCYTSS----- 54
Db 2 DDSPGPTSKARDKAENAENTSSDSDSEVSASEKSESRPSSEKKKVIITRIVPVRP 61
QY 55 -----FRSISTYSTPGRRKPRVHRPRSPRIIEKDIPLLEFPKSSDLMPVHEHIMVI 108
Db 62 PTRDKGHRVNLLESNESETKSLYQRAKEGIPSKGPEIKLPTTSEQYYDLSEVLN-- 119
QY 109 AIYEVLNFCVTLRLSPFERPEFCAALVSQEQTLMAEMHVLLKAVLREEDTNTTFGP 168
Db 120 -----PARMEG-----RELTNA----- 132
QY 169 ADLKDSYNSLYFIDGNTWPEVLRYCESDKYHHVLPYQEAEDYPYGPVENKIKVLQFL 228
Db 133 ---YDAVRNKNYVLP-----KSVCEAD-----LQKV 156
QY 229 VDQFLTNTARELMSEGVYQYDDH--CRVCHKGLDLLCETCSAVYHLECVKPPLEVP 286
Db 157 IGSEF-SCDVQCELIQSIIMTKGHRFCQOCILVAEVRSGNTCPTCRQLNGSKRELQODP 215
QY 287 E-DEMOCEVCVAHKVPGVTDCAVFIQKN-----KPYIRHEPIGYDRSRKRYWFLNRR 337
Db 216 RFQLIYQVVSISVG--RWMAENREHEKDVTFGRKGYIEG---GSDMNKRYGIDPNSK 270
QY 338 LIIEEDTENENKRI-WYSTKVLQALIEDLDKDYWEALCKILEMREIEIHRHMDITE 396
Db 271 LKAPRLKSAGRKKIRWFHES-----DED--GSVRKVMESKGA-----PKE 310
QY 397 DLTNKARGSNKSLAANEIILSIRAKGIDID---NVKSPETEKDKNETENDSKDAEK 453
Db 311 DFTNYLENDKEGTSVAAEKEVLE-----EGEMDFPIEIKSSDEQTLDDDEESMLDSDF 365
QY 454 NREEFEQSL-----EKSDDKTPDDPEQCKSEETEVGDKGNSVSANL 498

Db 366 EISDNEDVSKPSCSTSKKTTNRRSDSSSESDNRDNLQKKKKRM-----KKNVPKTD 419
QY 499 GDMTNTATSETPSEGRSPVGLSETPSSNMAEKVASELQDVPPEPNKTCESNTS 558
Db 420 GSDVSNESPEDASGE-----VVATKL-----IKESKKKCGRPKKK 456
QY 559 ATTTSIQPNLENSSSELNSSQSESAAAD-----DPENGERESHPTVPSIQE 606
Db 457 FAPELIEGDIPTPSEDSLTSSDEERDDNAADYAFVFKBFNRPD---RRDGH----- 506
QY 607 EIVGDTTSEKSTGELSPEGAGKAGSGSTRIITLRNPDKSLQSLKQAAAAHEANKL 666
Db 507 -----PEKD-----KLYNDFMIDM-----NHQVDRK 528
QY 667 FKEGKEVLYVNSQGEISRLSTKKEVIMKGINNYFKLQEGKYRVYHNOYSTNSFALNKH 726
Db 529 FEKDEGTHVSDSDNSSEHSEDEAE-----DRESSIDS----- 560
QY 727 QHREDHDKR--RHIAHKFCLTPAGEFKWGSVHGSKVLTTSTLTLTTQLENNIPSPFFH 784
Db 561 -----EHEKEISKFLSHRQPLNPNTSVDDDCQV-----ITVVKKD----- 595
QY 785 PNWASHRANWIKAVQMCSPREFALALALECAVKPVVMLPIMREFLGHTRLHRMTSIE 844
Db 596 -----VKOSATISKP-----GETSPDSSSKIE- 617
QY 845 EEKEVKKKKKEKEEETMQOATWVYTFPVKQHVQKQGEYRVTYGYSWISKTHVY 904
Db 618 EKPKDV--SEEVSDDEMTPEHITADKGTDTFLNIMEHDDM-----YGG-----YLF 663
QY 905 RFVPKLPNGTNNVNRKSLCTKNNMDENMDESRRKCSRPKIKIIEPD-----SEKDEVK 960
Db 664 R-----PGDGTGIS--RPKVQAPGNTRLSMNVCPVAVLKEGKKLVIPPEYISSDET 717
QY 961 GSDAAKAGADQNMEDISKITEKKDQDVKELSDSKCPKEPMEVDDDMKTES----- 1014
Db 718 LSDSEETSPAEQMEQSETSEAGPSTIIKTSGETRETQSSSPSEPTSRKMKHKLDT 777
QY 1015 -----VNCQESSQDVVVNVSEGFHLRT-----SYKKKTKSKLDGLERRIKQ 1057
Db 778 RRRKLADDSDLSDFDVSIDGNELVATGPKPIIKHVFDYDSANRMPKSNLFTGRRNARE 837
QY 1058 FTLEEKQRL-----EKILEGIGIKGISTNSTNNKLSNPVITKAKGCCQSDSRQOE 1110
Db 838 IPMEETSRLAEQVAHEEYKIHRRRQVYVLEAVEAASKLVNYVDVTTEEE-----IEEE 891
QY 1111 QSPNANNQDPELTIQSCSDSSVLRMSDPSHTT---NKLYPKDRVLDVDSIRSPETKC 1166
Db 892 ETPB-----EEVVK--VASPTAPIATENPTTSTAPFEEGVAMKETPIEIFF--DPDEPC 942
QY 1167 PKNSIENDI-----EEKVSDLASRGOEPTKSKTKGNDFFIDDSKLASADDIGTLIC 1218
Db 943 SSAQAQRELIIERVGKEQIIEDSLBQNRKPS-SKT-----VKESREAA----- 987
QY 1219 KKKPLIQESDRIIVSSSKALHSSVPKSTNDRDATPLSRAMDFEGKLGCDSESNSTLEN 1278
Db 988 --QEPRIKDEMESEQKKDADNPTVEVDKESASESSDKSDFE-----DETLD 1036
QY 1279 SSTVSTQDSSSEDIMVQNSNESISEOFTR-----EODVEVLEPLKCELVGSGSTGNC 1332
Db 1037 QSKTVKLSLKHETV-----SDEEI-EDFDTKFEFVATADAKMIKRTIGETVSTEFF- 1088
QY 1333 EDRLPVKGTEANGKPSQKKLEERP-----VNKSDQIKLNTTQKNNENRESEK-- 1384
Db 1089 -----KLVAQQPAVTDVLAALGFCVRNTDQEFSTIKETCKRTKNKPNDDSMV 1135
QY 1385 KGORTSTFQNGK-----DNKPKYIKYKCEGLKEISESRVSGNVPPKVNINIKIIPEN 1437
Db 1136 KNFRES-FAAKHRPVPRKPLTNTERMYI--ERAHMVYKHV--DMEPLHMKILIALQK 1190
QY 1438 DIKSLTVKES--AIRPFINGDVIMEDFNERNSETKSHLLSSSDABGNVRSLETLTLPST 1494
Db 1191 QIAATCANUSQPVTVTPTEEHAEQVLLHNLQNPSILRPLL-----NNPQFALTLLHKAQ 1243

Db	1002	QVTSETTTTTVTREYNDPEDEQEHQPHSPAPSSHTAEHPHIVETTTTTVTTRFQF	1061
QY	483	EPTEV-----GKGNSVSANLGN-----TTNATSEETSPSE--GRSPVGCLSTPDS	528
Db	1062	EPELEYKQEDNSRKSPPSHSQENLVETTTTTVTSEYDYDEPHEFEQETGKSPAPS	1121
QY	529	SNAEKKVASLPODVPENPKTCBSSNTSATTSTIQPN-LSNSN-----SSELNSS	580
Db	1122	HVESERQV---LESVPASDPRHVMTTTTTVTTRQPHDDSESRDNPNDVVEAQSIHSS	1178
QY	581	QSESAAADDPNGERESHPTVSIQEIY--GDFTEKSTGSELSESPGACKGASGSTRII	638
Db	1179	NTHSHQOFSEKDDSGSPVKSDEPVKHESYKFETSTTE-----TRF	1224
QY	639	TRLRNPDSLQSLKSOQVAAAHAANKLFKEGKVLVNSQGEISRLSTKKVIM-KGNI	697
Db	1225	DR-----PELDSERLSEPA-----QSPMEVSPITTEENIVRPSVV	1260
QY	698	NNYFKLGQEGKY-RVYHNOYSTNSFALNKHQHREHDKRHLAHKCLPAPGEFKWNGSV	756
Db	1261	KSEY--GSGRHVPSTIETTTTTVTREFYDDQDEHENQTSBELRSSIPTBEEHEGS-	1317
QY	757	HGSKVLTISTLRLTITOLENNIPSSF-----FHPNASHRANWKA-----VQMS	802
Db	1318	HLFKETTTTT--TVTRFYDEPENVEELODPQFSPAPSSHVESGIIHASESPVAQOQETP	1374
QY	803	KPREF-----PAAQYFHEEYESHVLTEQAPLLTEQHQHPESGDESOGELGSKVLGFAK	1434
Db	1375	QTRFHEPSPAQYFHEEYESHVLTEQAPLLTEQHQHPESGDESOGELGSKVLGFAK	1434
QY	820	P-----VVMPLWREFLG-----HTRLHRMTSIE-REEKEKVKKKKEEETMQOATW	868
Db	1435	AGVAGVGVAAAPVALAAVAKAAYDALKKDDDEDOEERESLLRQERSIDSPHASEBQS-	1492
QY	869	VXYTPPVKHQVKKQK--GEERYVTCYGGWSMISKTHVRYFVP-----KLPG--NTNVNR	919
Db	1493	QIEEHERFEESPVPSEKHVHTTTTTVTTRFYDFEHPVLSQEIEGKGPADSE	1549
QY	920	KSL-----BGTKNNDNDESKRKCSRPKKIKTEPDSEKDEKVGSDAAKAD-----Q	970
Db	1550	KSLPHVVETTTTTVTREFDKNDSE--SPVPSEKQEPTTVSREV--YETABGDEPHY	1605
QY	971	NEMDSKITEKDDQVKKLLSDSKPKCKEPEWY-----DDDKTE-----	1012
Db	1606	TETTTTTVTKEVIDDSQEM--GDDDEPKQESQVVTETTTTTMTSREYDNDDETRSEAGD	1663
QY	1013	SHV-----NCOESSQVDVY-----NVSEGFHLRTSYKKKTKSS	1045
Db	1664	SHITETKTTTVVREFHGGQPEETEETDEVELPPKIEEDNVSEYSESSTSVSREVRPD	1723
QY	1046	KLDGLLERRI-----KQFTLEEKQRLKTK	1070
Db	1724	E-PHIETTTTTVTREYHNEPEYDQKDAAPISFSQEHODDDSQASHDQHDRESVPE	1782
QY	1071	LEGGIKGICKTST-----NSSKNLSSSPVITKAKE--CCQSDSMRQEQSPNA-	1115
Db	1783	SEKSVKHTTTTTTTTTVTROLQYDDEASEIRGESPVATEBEHVESVTSKSDDESEHQVPSVI	1842
QY	1116	-----NNDQPEDLIOGCSQSDSSVLRMSDP-----SH-----TTN	1145
Db	1843	ETTTTTVTREYDQDE--LOREDHTQSEERSSIPTETETHEDSHLKEKTTTTTTVTTR	1900
QY	1146	KLYPDRVLDDYSIRPETKCPKQNSIENDIEBKVSDLASRGOEPTKSKTKGNDFIDDS	1205
Db	1901	EYDEPENVE--KLQDSQSLSPSSHVESIEIYVPESPVAKQEIPTRE-----PHDS	1952
QY	1206	KLAS---ADDITLCKNKKPLIQESDPIVSSSKSALHSSVPKSTNDRATPL-SRAMD	1261
Db	1953	PAAQYFHEDEYEHQVTEQAPLLTEQ-----HQ--PESGEESDGEFGSKVLG	1999
QY	1262	FEKGLGC-----DSESNSTL-----ENSDTVSIOD	1287
Db			
Db	2000	FAKKAGMWAGVVAAAPVALAAVAKAAYDALKKEDDEDOEERESLIREERSFDSPHASE	2059
QY	1288	SSEEDMIVONSNESISEQFTRREQDVEVLEPLKCELVSGES-TGNCE-----	1333
Db	2060	QSOIEKEHRFEESPVPSEKHHDOSSALPOESVSQPIEXESRTFNDESEFGVKS DHVTE	2119
QY	1334	DLPLVKGTTANQKKPSQ--OKKLEER---PVNKCSDQI-KLKNTT---DKKNENRESEK	1384
Db	2120	DOESLKSPPESGEAFQFTSEKQDRSDSPIHSOKEDIQFQNESSPEDEVKSEQPHDEEK	2179
QY	1385	KG-QRTSTQINGKONKPKIYLKGECLKEISERSVVSGNVEKVNINIIIPNDIKSLT	1443
Db	2180	PDLEROGSYSSGSPKSP-----GGSITGLDEKALSGVQEPEDR-----PENFAESHE	2228
QY	1444	VKESAIRPFIINGDVIIMEDFERNSSSTKSHL-----LSSDAEGNYRDSLETLPSTKES--	1497
Db	2229	KTEAT-----SDENLFESDKYAPASPVPSESDSSNRVETTTTTVTTRHFHE	2274
QY	1498	-----DSTQTTTPSACSPESNSVNOVEDMEIETSEVKVKVTSPITSEEP--SNLSN	1546
Db	2275	PEDDHSYVVESQOESYSSSGSPVPSEKSVDRV---IETTTTTVTTRHFHAEDEIPTIVES	2330
QY	1547	DFIDENGLPINKNENNVGESKRTVITEVTMTSTVATE-----SKTVLKVEKGDK	1597
Db	2331	SHDDQAASVPSEEDVHEQ-----IQTTTTTTVTKEHFVDPDDDEIDSEHMNESDKYAS	2383
QY	1598	QTVVSSSTENCAKSTVTTTTTTVTK-----LSTPSTGGSDII---SVKEQSK	1641
Db	2384	GSPVPSEEDSSRVETTTTTVTTRHFEPEDDHSPPVQTOEYSASESPVPSEKSVERVIE	2443
QY	1642	TVVTTTTVT-----DSLTTTGGTLVTSMTVSKE-----YSTRDKVKLKMFX	1680
Db	2444	TTTTTTVTTRHFEDHILGQOQESDQIPSEISITSENMDRETSSSPQSNRDEEVLPA	2503
QY	1681	SRPKK-----TRSGTALPSYRFVTKSKSI FVLPNDDOLKLRKGGTREVYPFN	1731
Db	2504	IAPYKOPTYGRVDSHDADAPSA-----PSAESPIQAYKQESBAHLSLEQFQQRSSVS	2557
QY	1732	YNAKPAALDW-----PYSPRPTFGITWRYRLQTVKSLAGVSLMLRLL-----	1774
Db	2558	HEDSPAAYFHDNESDHDSPVSDRAPLLTEQHQHPESGEESDGEFGSKVLGFAKKAGM	2617
QY	1775	-----WASLRWD-----DMAAKVPPGGGSTR-----TETSETE	1802
Db	2618	VAGGVVAAAPVALAAVAKAAYDALKKDDDEDENPDEQEKLLPKSPERQVLNVPESDSE	2677
QY	1803	ITTEIIRKRRDGPYGIREFEYCIKILICPIGVPEPKETP---TPORKGLRSSALRPKR	1859
Db	2678	ISEIE-LEYSPTSEKSESQCYTETVRTTTVTTRVREYLDPPQSVTRSRSPSEHDISEQAP	2736
QY	1860	ETPKQTPGVIIETWA-----EELELWEIRAFARVERKEKAQAVEQQAQR	1906
Db	2737	ESPVEQDPVVEKTTTVIRQYHDEPPQEIEEQIPEEVTVLREVYESPEGDEPEQH---Y	2793
QY	1907	LEQOKPTVATST-----TSPSTSTISIS---PAQKVMVAIPISGVITGT-----	1949
Db	2794	IETKTTTTITKEVHVPEEDVQISPVHSETSVSEKOLPADEQLEDEVTETTTATVTRER	2853
QY	1950	-----KWLTTTKVGSPTVT---FOQNKHFH-----	1973
Db	2854	YEEPEVRPPSGSEADDESHAPKYMETT---TTTTVTREYVESEDEHQOQVQDRDESP	2910
QY	1974	-----TFATWK-----QGOSSNGVVQOQVKLGIIPSSSTGTSTQFTTSPQR	2016
Db	2911	APSEDSVKHVIETTTTTVTTEERYEPEDSHSPVSEDDVHGFVKTTTTTTTTVTRHEFEPE	2970
QY	2017	TATVTIRPNTSGGGTTSNSQVITGQIRPGMTVIRTPLO-----QSTLGAIRTPV	2069
Db	2971	-----DPPSDHEHVVERYSRAGSPV---PSEEDSSREIETTTTTVTTRHF	3013
QY	2070	MVQPGAPQOVMQV--IIRGPVSTAVSAPNTVSSST-----PGKSLSATSTSN	2117
Db	3014	ELEDQOEHVVESEQEVSAGSPVPSEKSERVETTTTTVTTRHFHEDDIIPTIVETSHD	3073

Db 1640 ECTNLEGSFKCEHSGFEGDIGIKCTNPLERSCEDVEKFCGRVDHVSCLSVRYNGSLSS 1699
Qy 1102 -CGSD-SMRQEQSPNA-----NNQPEDLI----- 1124
Db 1700 VCECEPGFRFEKESNCVDIDECEESRNCDPASAVCVNTEGSYRCBAEGYEGGVCT 1759
Qy 1125 -----QCGSQSDSVLRM----- 1145
Db 1760 DIBECDRGAGCSMAMCINRMSCGCKCMAGYTGCGATCIKIEEPKSKDTACTDEWSR 1819
Qy 1146 --KLYPKDRVLDVDSIRSPETKCPKONSIEENDIEE-KVSDLASRQGE-----PT 1191
Db 1820 LCELEKKQCYDEEVEPQCGACLPGRHPINGTCOSLQISGLCAQKNDCKNKAECIDIHDP 1879
Qy 1192 KSKTKGNDFI-----DDSKLASADDICTLIC-----KNKKPLI 1225
Db 1880 SHFCSCPDGFIDGMCIDDVDECNNAGMCDDEKTCENTIGSFNCVCKLEGFKKVKDCV 1939
Qy 1226 QESDITVSSKSKALHSSVPKSTNDRDATPLSRAMDFEGKLGCDSESNSTLENSDTSVI 1285
Db 1940 DEKKQP--NREKIEIDENSNSNGQEKPTTKGI--VSSTSATSSSESTTAEPHVTTIS 1995
Qy 1286 QDSSEEDMIVQNESISEQFRREQDVEVLEPLKCELVSGESTGNCEDRLPKVKGTEANG 1345
Db 1996 STTSTKDMTSKSPENVMTSEPEVSTS-----SKSTTASETTVSTSTPSESS- 2044
Qy 1346 KKPQOQKKLRBP--VNKCDQIKLKNITDKNNENRESEKKQORTSTFOINGKDNKPKI 1403
Db 2045 ---SSEAPLTSPPATTEVITESSVKSTTPKESSSEITVKLSKS----- 2087
Qy 1404 YLAGECLKEISESRVSGNVBPV--NNINKIIPENDIKSLTVKESAIRPFINGDVIMED 1461
Db 2088 -----PEYTESVSKSPSTPTTSQSVTSTVPETSKSTVLSSEAPVTSTSTEV--- 2136
Qy 1462 FNERNSSETSHLLSSLDACGNRDSLETLPSTKESDSTOTT----- 1503
Db 2137 ---HTSETPPLSASSTT---GDTNSTPTSPSSLASVKSSTAPEGTSASVAPVKLSSL 2189
Qy 1504 TPSASCPESSVNOVEDMEIETSEVKVKTSPITSEESNLNDIFDENGPKINKENVN 1563
Db 2190 SPDVSQPSKTFDATESSTVQASSETSGTSKSTSEPESHVTKLSITSS---NPSSV 2245
Qy 1564 GESKRKIVITEVMTSTVATESKTVIKVEKGDKQTVVSTENCASKT-----VTITTTTV 1619
Db 2246 VTSKSTPTVPSTEQPTSTTPSGQSLTPMNSNEVLTTSEPHVLSLSPDVSQSSTTP 2305
Qy 1620 TKLS-----TPSTGGSDVIISVKEQSKTVVTTVTDTSLTTGTLVTMVSKEYSTRD 1673
Db 2306 NNLSSESTVETPKTSSEVLSNSEPSTEAPTLSPDILSTTTNNLSQSSTV-----STED 2361
Qy 1674 KVLMKFSRPPKTRSGTALPSYRKFTVKTKSIFVLPNDLKKLARKGGIRE-VPYFNY 1732
Db 2362 RSEISSENSEKPTSAPELVTSVTHVASSSPD---VPTESSEPDLLTGSSTENIP--EA 2415
Qy 1733 NAKPALDIWYPP-----SPRPTGIIWRYLQIVKSLAGVSLMLRLWLASLRDMDMAKV 1787
Db 2416 SSKQTSSTPTOTTASEPTKSTMSPLDSTTSNVLSSE-----ST 2458
Qy 1788 PPGGGSTRTTETSE-----ITTEIIRKRDVGYPYRIFCYCKIRKIICPGVPTPKETPT 1843
Db 2459 TPSSSKSPVSSSTEGISVVTSTFSK-----VPES----- 2489
Qy 1844 PQKGLRSSALRPPKPTPKQTPGVIIETWVAEELE-----LWEIRAFARV 1891
Db 2490 -----TSSVEEDLAKT--TPSPILEETTTASSETSEPLTDSLTVSRIHELTTSEN 2542
Qy 1892 EKEKAQAVEQAQKRLKQKPTVIATSTTSTSTSTISPAQKVWVAPISGVTGCTKM 1951
Db 2543 PRE-SESTTTSSSSKSPQEPAGILTSTVVVPTSSVSLITASE-----IEAITSNTPF 2594
Qy 1952 VLATKVGSPATVTFQKNKNHQTFAWVKOGQNSGVVQVKVGLIIPSTGTSTQSTFT 2011
Db 2595 -----KQGR-----PITTSFKSLVKST--TSPSTVT 2619

Qy 2012 SFOPRTAIVIRPNTSGGTTSSQVITQPIRGMVTIVNTPLQOSTLGAIIIRTPMV 2071
Db 2620 SSEPSESTKR-----TTVSTTVST-----TTPTTETTSESLLIT---A 2655
Qy 2072 QPGAPQOVMTQILRG--QPVSTAVSAPNTVSSTPGQKSLTSATSTSIQSSASQPPRPQ 2129
Db 2656 APSKPTSTSTESSEAPTTPAKTSTKPSNVSS-----TSRKSTENVETSTSO----- 2702
Qy 2130 GOVKLTMAQLTQTOGHGNOGLTVIQQGQTTGQLQLIPQGVTVLPQPGQQLMAQMP 2189
Db 2703 -----SGSLE-----STMS 2712
Qy 2190 NGTVQRELFPLATTATTATTTTTV-STTAAGTGEQRQSKLS---PMQVHQDKTLPPA 2245
Db 2713 STSEPETNPAVTVSSEASTTLEENSTSSPTSSSEASVKLSLFPESITSEAVTV--- 2769
Qy 2246 QSSSVGPAKQAPQTAQPSARPQOTQPSQAQPEVQTOPEVOTQTTVSSHVPSPAQTHA 2305
Db 2770 --SSRAPAEI---TMSSESHREITSVSSEPEIPLSTTVSPNVVTASSIPSE----- 2818
Qy 2306 QSSKPQVAAQSQP--QSNVQGSQSPVRQSPQSTRIRPSTPSQL-----SPGQSOVQTTTS 2359
Db 2819 -----EPILSSVTSSSPRV-----RLITGTPDDLIVSVTVPSHGNNRQNTA 2861
Qy 2360 QPIPIQHTSLQIPSQGQSQPQVQSSQTQTLSSGQTLNQVSVSSPSRPQLIQOQP 2417
Db 2862 SSVPSNSTSPILPSESLTTPQP-PTTTTAKPATTSGRKPPSIQPPAEMFTTPAP 2918

RESULT 7

A35085
C:Species: Drosophila melanogaster
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 20-Sep-1999
C:Accession: A35085; A38240
R:Mazo, A.M.; Huang, D.H.; Mozer, B.A.; David, I.B.
Proc. Natl. Acad. Sci. U.S.A. 87, 2112-2116, 1990
A:Title: The trithorax gene, a trans-acting regulator of the bithorax complex in Dros
A:Reference number: A35085; MUID:90192757; PMID:2107543
A:Accession: A35085
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3759 <RNA>
R:Mazo, A.M.
Submitted to GenBank, January 1990
A:Reference number: A38240
A:Accession: A38240
A:Molecule type: mRNA
A:Residues: 1-2361, 'Y', 2363-2397, 'N', 2399-2405, 'N', 2407-2411, 'N', 2413-3759 <MA2>
A:Cross-references: GB:M31617; NID:g158817; PID:g158818
C:Genetics:
A:Gene: FlyBase:trx
A:Cross-references: FlyBase:FBgn0003862
A:Superfamily: Drosophila trithorax protein
C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 3.1%; Score 462.5; DB 2; Length 3759;
Best Local Similarity 18.7%; Pred. No. 3.2e-09;
Matches 627; Conservative 422; Mismatches 1151; Indels 1157; Gaps 156;
Qy 5 EEEEDGDAEETQSDDD-----EEDMEEDDDSDYPMEEDDDDDDDASYCTESSFRSHST 60
Db 538 DEDEGGVTFRNDSPEDONNAEDDEMDDDDDDEAEEDQDNEEDNDDEAAASEKSAETKS 597
Qy 61 YSS-----TPGR--RKPRVHRPSPILEEKDIPPLEFKPSSEDLMPN 101
Db 598 AGADERDPDEKQLVMDSHFVLPKRKRSRIKPKRLLEGAT-----STKKPLSLGD 651
Qy 102 EHTMNVIAI-----YEVLRNFGT-----VLRLSPRFE 129
Db 652 SKGNVFGTSSSSAGSTASTFSASTNLKLGKEFFNGLTKPNSSAAGIFVLQRPQRFQ 711

Db 722 ADTTQIETETVVDDSKKKVLKKK--TEKDSF-----ISQSETPPVPEP 766
Qy 777 NIPSSFFHPNASHRANWIKAVQWCKPREFALAL-AILECAVPPV--MLPTWREFLGH 833
Db 767 TKP-----ASEAQKIAEVNKAQKQEVDDNLAKRAEVAAKIADEKLEIAE- 814
Qy 834 TRLHMTSIEBEEKEKVKKEKQEBEETMQQTWVYTFPPVKHVMKQKGEERYVTGYG 893
Db 815 ANIKKTAEEVAARKQEKDEQLKLETEVSVSKSAEAKLELEKQAOI--KKAAB- 865
Qy 894 GWSWISKTHYRVPKLPGLNTVNVYRKSLEGTNNMDNM----- 933
Db 866 -----ADAVKKOKELNEKNLEAAKSAADKLLEESAASKVKYSESVKF 912
Qy 934 DESDKRCSRPKIKTEPSEKD-EVKGSDAAKADQ--NEMDISKITEKKDODKELL 990
Db 913 GEEKKTAGEKTVOVESEPTSKKTIIDTKDVGATEPADETPKKIILKKTEKSDSSISQKS 972
Qy 991 DSDSDPKCK---EEP-----MEVDDDMKTESHV-----NCQES 1020
Db 973 ATDEKYSKQKQODEPTKPAVSETQMVTEADKSKQKETDEKLKLDAAETAAKTQEADEK 1032
Qy 1021 SOVD-----VNVSE-----GPHLRTSYKKTKSSKLDGL-LERRIK-----OF 1058
Db 1033 SKLDAQEKIKVSEDDAARKEKELNDKLESEIATKKAADKLEBQAKAAAEVEA 1092
Qy 1059 TLEEKQREKIKLEGGIKGIGKSTNSSKNLSESPVITK---AKEQCSDSMRQEQSPNA 1115
Db 1093 AKKQEKDEQKLD-----TEAASKAAAEKLELEKQAOIKKAAGDAVKKQKELDE 1144
Qy 1116 NNDQPELIIQCCSDSSVLRMSPSHITNKLPKDRVLDVDSIRSPETKCPKQNSIEND 1175
Db 1145 KNK-----LEANKKSAAGLKIEESAASKQVEEQAKLDAQTKAKTAB--KOTKLEKD 1197
Qy 1176 IEKVSDLASRGQPTKSKTGNDFDIDSKLASADDIGTLICKNKKPLIQEESDRTVS- 1234
Db 1198 -EKSTKESKETEYDEPKKK-----VLKKK-----TEKSDSSISQ 1232
Qy 1235 ---SSKALHSVPKSTNDR---DATPLSRAMDFEGKLGCD-----SESNSTL 1276
Db 1233 KSETSKTVESAGSESETOKVADAARKQETDEKQKLEAEITAKKSADEKSKLEAESKL 1292
Qy 1277 ENSSDTVIODSSSEDMIVONSIESIQFRTQEDVEVLEPLKCELVSGESTGNCEDRL 1336
Db 1293 KAAAEVAARKQEKDEQLKLDTEAASKAAAEKLELEKQSHIK----- 1336
Qy 1337 PVKTEANGKKPSQOKKLEER-----PVNKCSDQIKLNTTDKNN-----NENRESE 1383
Db 1337 --KAAEVDVAVK--ROKELEEKORLESEAAATKKADEKLEKLEEQKKAETALIEIOKEQ 1392
Qy 1384 KKGQRTSTFQINGKDNPKIYLKGECL-LKEISESRVVGNVPEKPVNNINKIIPENDIKSL 1442
Db 1393 KLAQEQSLDEAKKSAEKQKLESETSKOTEEAPKESVDEKPKKVLKKKTEKSD--SSI 1451
Qy 1443 TVKESAIRPFTN-GDVIMEDFN-----ERNSETKSHLSSSDABEGNTRDSLETL 1491
Db 1452 SOKSKASAKTVDAAEFTLESDFNLVEKKTQVKVEQSPDETSATIKRDPAQ-----KTEEI 1506
Qy 1492 PSTKESDSTQTTTSPASCPESNSVQVEDMEIETSE--VKKVT--SSPITSE-----EESN 1543
Db 1507 SKQDDGDEKKTU-----DGKPPKPEDSEATPKKRVVKKTKQKSDSVASDASLADYSK 1559
Qy 1544 LSNDFIDENGLPINKNENVNESKRKTVITEVTMTSTVATESKTVIKVEKGDK----- 1597
Db 1560 LSDDVEEPKPKVLKKKT-----EKSDSVISE-TSSVDTIKPESE-VEIPTKAEQMIHLNR 1613
Qy 1598 ----QTVVSSSTENCAK-STVTTTTTKLSTPSTGGSDIISVKEQSKTVTTTT----- 1647
Db 1614 FSTDSAVESEPKNAHKDTEKTTDDMTTRKSSA-----IFSDEQSISSKTSSEGRRR 1667
Qy 1648 -----VTDLSLTTTGGT-----LVTSMTVSKEYSTRDK-----VKL 1677
Db 1668 RRTGFAKFAKSDTLALRGDNVEIEAEALLAEDDVTWVNGKDADLNSRCHESHHTFTFT 1727

Qy 1678 MFSRPRKTRSGTALPSYRKFEVTKSTKKSIFVLNDDLLKILARKGGIRE-----VPYFN 1731
Db 1728 LIIDEVEPTDSOMEITATCGTESHTTLKVEELPVDFVKYLPKRKTSKGEQEVITSITLN 1787
Qy 1732 YNAKPAIDIWPPSPRPRTFGITWRYRLQTVKSLAGVSLMLRLLLWASLWDDMAA-KVPPG 1790
Db 1788 HPIDISKVW-LKDGP-LEINKDYSIDTVG--CSVSLTLR-----RAKIEDSGKYKVVCD 1839
Qy 1791 GSGRT-----ETSETEITTT-----EIIKRRDVGVP-----GIRFEYC 1824
Db 1840 GUDCSTHLSIOGKPVLKNNVSETKPVITVDKDDQDQSLVAYDSNPEASFMTVDGKLEFD 1899
Qy 1825 IRKIICPI-----GVPEPK---ETPTPQRKGLRSSALRPRPTPTKOTGPVIE- 1871
Db 1900 GRSRIDVDDGLKTRKGVSKTDAGEYEVKLNKNEFEVAQKFDVKVNDTPSAPGDVSVMK 1959
Qy 1872 -----TWAEELWELEWRAFAERVEKEKA-----OAVEQQAOKR-----LEOK 1911
Db 1960 AESDCLHIEWTAPTEGNGAEVTSYV--LEKRESGRKKFKHVKATVNGKKTSTVDDLEIET 2017
Qy 1912 PTVIATSTTSP-----TSSTTSTISPAQ--KVMVAPISGVT----- 1947
Db 2018 PIVIRIAAVNKFGTGEFETKPVQSGSPFQVPTVEFPTIDNVTSTSCSLSWPKPIEDGG 2077
Qy 1948 -----GTMVLTKVGSFATVTFQONKNEFHOTFATWVKQ-OSN 1985
Db 2078 SPVYGYDVYKRENEGEWQKMGNEELVFTSEFNVRA--LSSGKEYEFKIEACNEAGLRN 2134
Qy 1986 SGVVOVQOKVLGIIPS-----STGTSQOFTTSFQPTATVIRPNTSG----- 2028
Db 2135 SNVSKKLTVGLVPEIILDMPMVKVLDNDKVEVTKSDEGEFFVQVYKSDGSSIASVD 2194
Qy 2029 SGGTTSNSO-----VITG-----POIRPG 2047
Db 2195 IGGPRSESAATSKCIIDGLREGIPYVFRVAARNHQTGEFSEPTIPVVVLADDAAPRLKA 2254
Qy 2048 MTVIRTPLOOSTL-----CKAIIRT---PVMVQGAPOQVMTQIIRG 2086
Db 2255 IKPVKIPKGELELRECHAAGHAPAEYIWKDGEIIPDENTEIVNEGSMALIIHELAG 2314
Qy 2087 QPVST-AVSAPNTVSTPGOKSLTSATSTNSIQSSASOPPRPOQG-QVKLT----- 2135
Db 2315 EDVGLYKVLVNIHGTAESEAVGISDVRAHNSFSSELTETIEEGHDIETCEVSEDAV 2374
Qy 2136 -----MAQLTQL-----TQHGNGOGLTV----- 2155
Db 2375 VNMVKGKKLVASDRVQFYAMARKRTLRIGKSTDAADSGVYKCTETDGRSRTGEVIVNEQ 2434
Qy 2156 -----IQOGQOT-----TGOLILPOGVTVLPGPGQOLMOAAMPNGTV 2193
Db 2435 EPHILVGPDAIVKOFGETMVLFCETSKPVKVKFKNGVEIWQPMKNAIMENDCKRATL 2494
Qy 2194 QRLFTPLATTATTASTTTTSTVTAAGTGEOROKLSPO-----MOVHODKTLPPAOS 2248
Db 2495 EIKNEFDKHDIGAYTASVSEKETSAKLVFEVAPNLIIPTEIRDGTVTHAGNEFDFA--- 2551
Qy 2249 SVGPAKAQPTAQPSPAPQPTQPOSPAQPEVOTQPTVTSVSHVSEAPQTHAOS 2308
Db 2552 -----VEFSGFPIPTHLLNNGTPKAIADVTEYDSDSVSRM---KDVTLDSNG 2597
Qy 2309 KPOVAAQSQPSNVQSQSPRVQSPQTRIRSTPSQLSPGQSQVQTTTSQPIQPHIT 2368
Db 2598 TVRVIAES-PLGOCIKEIPLKIID-----KPSAPCDL-----QFREVTEDSVFLSWP--- 2644
Qy 2369 SLOIPSQGQPOSPQVQSSQTQLSSQTLNOVSVSPS-RPQOLOIQOQOP----- 2417
Db 2645 --PLETNGAP-----LTVYIERKAVDNNRWRPCQGVKPHKLTFAEDLFCN 2689
Qy 2418 -----QVIAPQLOQO-----VOVLSIQSQVVAQIAQSQGVPOQIKLQPLQIOQSSA 2467
Db 2690 QVYGFRLIAVNEVGESEPCDITVDVLTLESSEPVSS--ESSELFVFKIALLRTP---QVTVA 2744

Db 1317 FMESPKQILDAAALSGKRLRTPKSKSEVPEDLAGIELFQTPSHKESMTNEKTKV 1376
QY 1707 IFVLPNDDL-----KKLARKGGIREVPYFNNAKPALEDIWPSPRPTFGITWR 1755
Db 1377 SYRASQPLVDTPSSKQPPKSLRKADTEE-----EFLAFKQTPSAG---- 1420
QY 1756 YRLQVTKSLAGVSLMLR-LLWASLRWDDMAAKVPGGSGSTRTESEITTEILIKRDV 1814
Db 1421 KAMHTPKPAVGEKDINTFLGTPVQKLDQPGNLP--GSNNRLQTRKEKAQALELT--- 1474
QY 1815 GPYGLR-----FVCRKIKIC--PIGV-ETPKETPTPORKGLRSS----- 1852
Db 1475 ---GREFLPQPCDNTADKTKKILCKSPQSDPADTPNTKQPKRSLKADVEEF 1531
QY 1853 -ALRKRKP-----ETPKQTPGVIIETWVAEELELWEIRAF-----AERVE- 1892
Db 1532 LAFRKLTPSAGKAMHTPKAA-----VGEK-----DINTFVGTPEVKLDDLGNLPGSK 1579
QY 1893 -----KEKAQAVEQAKRLEQKPTVIATSTTSPTSSSTTSPAKQVMVAPISGVT 1946
Db 1580 RRPQPKKAKALEDLAGFKELFQTPGHTESMTDDKITEVSCSPQDPVKTP- 1633
QY 1947 TGTVMVLTTKVGPATVTFQMKNFHOTFATVWQGSNGVVOQKVL--GIIPSSTG 2004
Db 1634 TSSKQRLKLSLGK-----VGVKEEVLVPGKLTQTS 1664
QY 2005 TSCQFTSFQPRATVTRPNTSGSGGT-----TSNSQVI-----TGQPIRPGMTVIRT 2053
Db 1665 KTTQT-----HRETAGDGKSIKAFKESAKQMLDPANVTGTMERWP-----RT 1706
QY 2054 P-----LQOSTLGRKAIIRTPVMQPGAPQVMTQIIRGQPVSTAVSANTVSTSPGQKS 2107
Db 1707 PKEEAQSLDLAGFKELFQTPDHTESITDDKTKI-----ACKSPPEES 1751
QY 2108 LYSATSTNIQSSASQPPRQGOVKLT--MAQLFOLTQGHGNGOGLVFIWQGOQTGQ 2165
Db 1752 MDTPST-----RRRPKTPGKRDIVEELSALKQLTQ-----TTH 1786
QY 2166 LQILIP-----OGVTLPGPQOQLM-----QAAMPNGHVQ-----RFLF-TPL 2201
Db 1787 TDKVPGEDKGINFRETAKQLDPAASVTGSKRQPTPKGAQPLEDLAGLKFQTPV 1846
QY 2202 ATATTASTTTTSTTA-----AGTGE--QRQSKLS-PQMVQHDQKTLPPAQSSSVGPA 2253
Db 1847 CTDKPTTHTKTIACRSQPDVPGTPTIFKQSKRSLRKADVEESLALRKRTPSVGKA 1906
QY 2254 KAQOTA-----QPSARPOQTPQSPAPQPEVQTPQVOTQTVSS 2294
Db 1907 MDTPKPAAGDEKDKAFNGTVPQKLDLPGNLFGSKRWQPTPKE-----KAQALEDLAG 1959
QY 2295 -----HVPSRAQPTHAQSSKPOVAQSQSNVQGSQPVVQSPQSTRIRPS----- 2341
Db 1960 FKELFQTPGTDKPT-TDEKTKIACKS-PQ-----PDPVDTPAKQRPKRLKADV 2010
QY 2342 -----TPS-----QLSPGQSQVOTTTSPIDPIQPHSL-QIP-SQGQSQSQP- 2382
Db 2011 EEEFLALRKRTPSAGKAMDTPKPAVSDEKNINTFVETPVQKLDLGNLFGSKRQPTPKE 2070
QY 2383 -----QVQSTQTLSSGQTLNGVSVSSPSRPOLOIQOQPOQVIAVP----- 2423
Db 2071 KABALEDLVGFKELFQTPGHTESMTDDKITEVSCSPQEPFKTSRSKQRLKIPLVKV 2130
QY 2424 QLOQGVQLSQI--QSQVVAQIAQSQGVQPIKL--QLPIQIQSSAVYTHQIQNVIV 2479
Db 2131 DMKEEPLAVSKLRTSETTQTHTEPTGDSKSIKAFKESPKQILDPA-----SVTG 2182
QY 2480 QASVQBLQVQOQLRDQKQKK-----QQQIEINVT-----PSKLLIKVEIIQ 2524
Db 2183 SRRLRTRKEKARALEDLVDFKELFAPGHTESMTIDKNTIPCKSPPELTDATSTK 2242
QY 2525 -----QVVMKHNAVIEHLKQKSWT-----PAERENORMVNCVMKYLIDKIDK 2571

2243 RCPKTRPRKEVKEELSAVERLQTSGQSTHTHKEPASGDGKVL----- 2287
QY 2572 EEQQAOKKRKRESEVQKSKQNA-----TKLSALLFKHKEQLRA----- 2611
Db 2288 --KORAKKPNPVEEPSRRRPRAPKEKAQPLEDLAGFTLETSHTOESLTAGKATKI 2345
QY 2612 -----EILKKRALDKDQIEVQBELKRDLIKKEKDLMLQAA----- 2649
Db 2346 PCBSPPLEVVDTTASTKRLHTRVQK-----VQVKEPSAVKFTQTSGETTDADKEPAGE 2400
QY 2650 -----ATAVAPCPPTVPLPAPAPPSPPP-----PPPGVQHGLLSTPPL 2691
Db 2401 DKGIKALKESAKOTPAASVTGRRRPRAPRESQAIEDLAGFKDPAAGHTESMTDDK 2460
QY 2692 PVASQKREKEEDSSSKKKMISTTSKETKDKLYCICKTIPYDESKFYIGCDRCQN 2751
Db 2461 TTKIPCKSSPELEDATATSSRRRPRTRAQKVEKEE--LLAVGK----- 2501
QY 2752 WYHGRCVGILOSEAELEIDYVCPQOSTEDAMTVLPLTEKYEGLKRVLSLQAHMAW 2811
Db 2502 -----LTQTSGE-----TTHTDKEVPVGEKGTKA----- 2525
QY 2812 PFLEVPDPN-DAPDYGVVKEPM-----DLATMEERVQR-RYEKLTIEFVAD 2856
Db 2526 -FKOPAKRNVDAEDVIGRRQPRAPKEKAQPLEDLASQFELSQTGTEELANGAD 2581
RESULT 10
F90073
hypoetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F90073
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cul, L.; O
ma, A.; Mizutani-Oi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Status: preliminary
A:Accession: F90073
A:Molecule type: DNA
A:Residues: 1-2271 <KUR>
A:Cross-references: GB:BA00018; PID:q13702612; PIDN:BA843752.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2447
Query Match 2.9%; Score 440; DB 2; Length 2271;
Best Local Similarity 18.3%; Pred. No. 1.2e-08;
Matches 442; Conservative 332; Mismatches 1000; Indels 642; Gaps 84;
QY 411 AAANEELIESIRAKGDIIDNVKSPETEKKNETENSKDAEKREFEDESLEKSDDK 470
Db 90 AASDAPLTSELNTOSETVGNQNS---TTIEASTADSTSVTKN-----SSSVQTSNSDT 141
QY 471 TPDDPDEGKSEEPVEVCDKGNVSANLGDNTTATSE---TSPSGRSPVGCLSRTP 526
Db 142 VSSE-----KSEKVT-----STTNSNQOKLSTSESTSS---KNTT 177
QY 527 DSSNMAEKVASELPQDVPENPKTCESNTSATTTSTQPNLENSNSSELSN----- 579
Db 178 SSSDTKSVASTSTEQPINTSTNQSTASNNSTQSTTSSVNLNKTSTTSTAPVKLRTF 237
QY 580 ----SOGESAKAADDPENGERSHTVPSIOEIV-----GDTSEKSTG--ELS 622
Db 238 SRLAMSTFASAA-----TTTAVTANTITVKNLKNQYMTSGNATYQOOSTGIVTLT 288
QY 623 ESPGAKGA-SGSTRIITLRNPDSKLSQKSOQVAAAHAAANKLFKEGK-----EVLVYNS 678
Db 289 QDAYSKGAIITLGTRI-----DSN-----KSFHSGKVLNKGKVEGHGNGDGIGFAFS 337
QY 679 QGEISRLSTKKEVIMKGINNINNYFKLGQEGKYRVYHNOYSTNSFA----- 722

cell proliferation antigen Ki-67, long form - human

C:Species: Homo sapiens (man)

C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999

C:Accession: A48666

R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde

J. Cell Biol. 123, 513-522, 1993

A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiquitous

ins.

A:Reference number: A48666; MUID:94043435; PMID:8227122

A:Accession: A48666

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-3256 <SCH>

C:Cross-references: EMBL:X65550; NID:g415818; PID:CAA46519.1; PID:g415819

C:Superfamily: kinase interaction domain homology

C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat

F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 2.9%; Score 439.5; DB 2; Length 3256;

Best Local Similarity 18.3%; Pred. No. 1.9e-08;

Matches 579; Conservative 432; Mismatches 1103; Indels 1049; Gaps 148;

300 VGVVDCVAEIKKPKYIRHEPIGYDRSRKRYWFLNRLIIEEDPTENENKKIWT----- 354

221 VP-TTQCLDNKKNE-----SPFWKLYESVKKELDVKSOKENVLQYCRKSG 265

355 -----YSTKVQVLAELIDCLDKOYWEAELCKIILEEMREEIHRHMDIPEDLTNKAAGSNKSF 410

266 LQTDVATKESADGLQ-----GETQLL-----VSRKSRPKSGSGHAVA 304

411 AAAN-EEILESIRAKGIDIDNVKSPET-----EKDKNET-----ENDSKDAENR 455

305 EPASPEQELDNQNGRVDVESQTPSKAVGASFPLYEPAKMKTPVOYSSQOONSPOKHKN 364

456 EEF-----EDQSLKLD-----SDKTPDDDDPECKSEETEVGDKNSV--ANLGDNTN 504

365 DLYTTGRRESVNLGSEFGKAGDKLTLPKLSRNRPTAKVEDAADATKLENLSKTRG 424

505 A-TSEETSPSE---GRSPVGCSET-----PDSSNNAEK-----KVASELPQDVPE 546

425 SIPTDVEVLPTETEHNEPFLTLWLTVQVERKIQKDSLKPEKLGTTAGQMCGLPLSSV 484

547 ENKTCESNNTS-----ATTTSIOPNL--ENSNSSSLNSSQSESASAA----- 588

485 DINFGDSINESGIPLKRVRVFGGHLRPELFDENLPNTPLKRGAPTRKKSILVMHTP 544

589 -----DDPE-NGERES---HTPVSTQIEIVG-----DFTSEKSTGELSESPG 626

545 PVLKKIIEQOPQPSCKQESGSEIHVEVKAQSLVSPAPSPRKTVPASDQRRRSCKTAPA 604

627 AGKAS-----GSTRITRLRNPD-----KLSQKSOQVAAAHAANKLFKEGKEVL 674

605 SSSKSOQTEVPKRGGERVATCLQKRVISRSQSHDILQMICSKRRSGASEAN-----LI 656

675 VVNSQGEISRLSTKK---EVIMKG---NINNYFKLGQBECKYRV--YHNOYSTN----- 720

657 VAKSWADVVLGAKQOTQTKVIKHGPQSRNKKQRAPATPKKPVGEVHSOFSTGHANSPECT 716

721 FALNKHQHRREDHDKRR--HLAHLFCGLTPAGEFK-----WNGSVHGSCKVLTITSLRL 769

717 IILGRAHTEKVVHVPARYVLANFISNQMDKEDLSGIAEMFKTPVKEQPOLT-STCHI 775

770 TITQLENNTPSSF-----FHPNASHRANWIKAVOMCSK--PREFALALALECAVK 819

776 AISNSENLIGKQFGQTDGSEELPLTSESFGNFFSAQNAAKQPSD-----KCSAS 827

820 PVVMLPIWREFLIGHT-----RLHRWTSIERE-----EKEK 849

828 P-----PLRQCQIRENGNNAKTPRNTYKMTSLETKTSDETETPSKTVSTVNRSGRSTEFNR 883

850 VKK--KEKKQEEETM-----QQAATWYKTFPVKHQVWQKGEYRVYTGGSWMI 898

884 IQKLVPESKEBTNTEIVEICILKRGOKATLLQO-----RRGEMKEI----- 925

899 SKTHYRVFVKPLPGNTVNYRKSLSEGTKNNDENMDESCK-----RKCS--RSPKKIKIE 951

926 -----ERPEET---YKENITELKEN--DEKMKAMKRSTRTGWQRCAPMSDLTLKSL 970

952 PDSE--KDEVKG-----SDAAKADQNQEMDISKITEKKQDVVVELLDSDSKCKE-EP 1002

971 PDTLMKDTARGQNLLOTDHAKAPKSEKGIITKM-----PCQSLQP 1012

1003 MEVDDDMKTESHVNCQESSQVDVNVSE-----GFHLRITS-----Y 1038

1013 -----EPINTPTHKQQLKASLGKVGKVELLAVGKFTRTSGTTHTHREPAGDKSIRTF 1068

1039 KKKTK-----SSKLDGLLERIKQFTLEEKORLEKIKLEGKIGKIGKIGTSTNSSLNLSBP 1093

1069 KESPKIILDPAAVVTGM--KKVPRTPKEEAQSLDL---AGFKELFOTPGPSEESMTDEK 1123

1094 VITRAKEGCSQDSMRQEOPNANDQPE-DLIOGCSQSDSVLRMSDPSHTTNKLYPKDR 1152

1124 T---TKIACKSPPPESVDPTPTSTQWPKRSRKADVEEFLALRLKLPISAGAKMLTPKPA 1180

1153 VLDDVSI-----RSPETKCPKONSIE-----DIEEKVSDL 1183

1181 GGDEKDIKAFMGTPVQKLDLAGTLPQSKRQLQTPKRAQALDLAGFKELFOTPGHTEEL 1240

1184 ASRGO-----EPTKSKTKGNDFIDDSKLASADDIGILI--CKNKKPLIQE 1227

1241 VAAGKTTKIPCDSPQSDPVDPTSTQKQ-----PKRSIRKADVEGELLACRLNMP----- 1290

1238 ESDTIVSSSKSALHSSVPKSTNDRD-----ATPLSRAMDFEGKLGCDSESNSTLSSSDT 1282

1291 -----SAGAMHTPKESVGEKDIIFVGTVPQKL-----DLTENLTGS 1329

1283 VSIQDSEEDMIVQNSNESISEQFRTREQDVVEVLEPLKCELVGSESTG--NCEDRLPVKGT 1341

1330 KRRPQTPKEAQALEDLTGFKELFOTPGHTEEA-----VAAKTTKMPCESSPPESAD 1382

1342 EANGKKPSQOKKLEERPYNKCSQDIKLKNTYDKNNENRESEKKGQRTSTQINGKDNKP 1401

1383 TPTSTRPQKPLEKRDVOKELSAUK-----KLTQTSGETTHDKVPGEGDS 1430

1402 KIYLGECLEKISESRVYSGNV--EPKVNINIKIIPENDIKSLTVKESAIRPFINGDVIME 1460

1431 INAFETAKQKLDPAASVTGSKRHPKTK--EKAQPLEDLAGW--KELFQTP-----VCT 1480

1461 DFNERNSSETKSHLLSSSDAEGNYRDSLETLPSTKESDSTQTTTPSASCPSNSVQVED 1520

1481 DKPTTHETTKIACRSQPD-----PVDTPSSKPSQSKSLRKVD 1519

1521 MEIETSEVKVYTS-----PITSSESNLNDIFIDENGLPINK--NENVNGESKR 1568

1520 VEEEFPAUKRTPSAGKAMHTPKPAVSGEKNIYA--FM---GTPVQKLDLTENLTGSKRR 1574

1569 KTVITEVTMTSTVA-----TESKTVIKVERGDKQTVVSVSSSENCAKSTVTVT 1616

1575 LQTPKEAQALEDLAGFKELFOTRGHTEESMT---NDKTAACKSSQPDLDKNPASSKR 1631

1617 TTVTKLSTPSPGGSDIISVKEQSKTVVTTTVDLSLTGTTGTLVTSMTVSKYSTRDKVK 1676

1632 RLKTSLG--KVGKVELLAVGLKLTQTSGETTHETPTGDKGSKMAFSPKQIILDSAS 1689

1677 LMKFSRPPKTRSGTA-----LPSYRK--FVTKSTKKSIFVLPNDLL----- 1715

1690 LTGSKRQLRTPKGSKEVPEDLAGFTLQFQTPSHTKESMTNEKTKVSVRASQPDLDVDTPT 1749

1716 -----KKLARKGGIREVPYFYNNAKPAALDWPYSPRPTFGITWRYLRQTVKSLAGVSL 1769

1750 SSKPQKSLRKADTEE-----EFLAFKQTPSAG-----KAMHTPKPAVGEEK 1793

1770 MLR--LLMASLRWDDMAAKVPPGGSTRTSTSTETITTTTEIKRRDVGPIGR----- 1820

1794 DINTPLGTVPQKLDQPGNLP--GSNRRLOTRKEKAQALELT-----GFRELFTPTCT 1844

Qy 1821 -----FEYCIRKIIC--PIGVP-ETPKETPTPQKGLRSS-----ALRPKRP----- 1859
Db 1845 DNPTADEKTKKILKSPQSDPADTPTNTKQRPKSLKADVEEFALFRKLTPSAGKAM 1904
Qy 1860 ETPKQTPVITETWVAEELELEWIRAF-----AEERVE-----KEKAQAVE 1900
Db 1905 HTPKAA-----VGEK-----DINTFVGPVEKLDLGNLPGSKRRPQTPKREKAKALE 1952
Qy 1901 QOAKRLEQOKPTVIATSTTSSTTSIPAKQVMVAPISGVSVTGTRKVLTKTKYGPSP 1960
Db 1953 DLAGEFKELFQTPGHTESMTDDKITEVCSKSPQDPVKTP-----TSSKORLKISLGK- 2005
Qy 1961 ATVTFQKNFHFQTFATWVKQGSNGVVOQKVL--GIIPSTGTSQOQTFSTFQPRTA 2018
Db 2006 -----VGKVEVLPVGLTKTQTSKTTQT----- 2028
Qy 2019 TVTIRPNTSGSGGT-----TSNSQVI-----TGQIRPGMTVIRTP-----LQOSTLG 2061
Db 2029 -----HRETAGDGSIKAFKESAKQMLDPANYGTMERWP-----RTPKEAQSLDLAGF 2079
Qy 2062 KAIRTPVMVQPAQVQVMTQIIRGQPVSTAVSAPNTVSTPGQKSLTSATSTSNIOSSA 2121
Db 2080 KELFQTPDHTDEESTDDKTKTKI-----ACKSPPPESMDTPTST----- 2117
Qy 2122 SQPRPOOGQVKLT--NAQLTOLTQGHGNOGLTVVIOGOGTTCQLQILP-----QGVTV 2175
Db 2118 RRRPKTPLGRDRDIVEELSALKQLTQ-----THTDKVPGEDKGINV 2159
Qy 2176 LPGAQOQLM-----QAAMPNGTVQ-----RFLF-TPLATTATTASTTTTV 2215
Db 2160 FRETAKQLDPAASVTSKQKQTPKGAQPLEDLAGLKELFQTPVCTDKPTTHEKTKTKI 2219
Qy 2216 STTA-----AGTGB--QRQSKLS--PQMVHQDKTLPPAQSSSVGPAKAQPOTA----- 2260
Db 2220 ACRSQPDVPGTPTIFRQSKRSRKADVEESLALRKRTPSVSGKAMDTKPKAGDEKDM 2279
Qy 2261 -----QPSARPQQTQPSQAQPEVQTPQVQTTVSS-----HVPSEAQPT 2303
Db 2280 KAFMGTPVQKLDLPNGNLPGSKRPQTPKE-----KAQLEDLAGFKELFQTPGTDKPT 2332
Qy 2304 HAQSSKQVAAQSQPQSVQSVQSPVRSQTSRIRPS-----TPS- 2344
Db 2333 -TDEKTKTKIACKS-PQ-----PDVDTASTKQRPKNLRKADVEEFALRKRTPSA 2383
Qy 2345 -----QLSPGQSQGVQTTTSOPIPIQHTSL-QIP-SQGOQSQSQP----- 2382
Db 2384 GKAMDTPKPAVSDKKNINTEFVTPVQKLDLGNLPGSKRQTPPKKAEALDLVGFKEL 2443
Qy 2383 -QVQSSQTLLSGQTLNOVSVSSPSRPLQIQPOQVIAVP-----QLQOQVQLSQI-- 2435
Db 2444 FQTPGHTESMTDDKITEVCSKSPQSPESFKTSRSSKORLKIPLVKVDKKEPLAVSKLTR 2503
Qy 2436 QSQVVAQIQAQSGVPOQIKL--OLPTIQOQSSAVQTHQIONVVTVQAASVQEQOLRVQ 2493
Db 2504 TSGETTQTHTEPTGDSKSIKAFKESPKIILDPAA-----SVTGSRRQLRTRKAKARA 2555
Qy 2494 LRQOQOKKK-----QOQIEINVNT-----PSKLLIKVELIQ-----KQVMKH 2531
Db 2556 LEDLVDFKELFSAPGHTESMTIDKNTKIPCKSPPELTDATSTKRCPTKTRPKRKEVKEE 2615
Qy 2532 NAVIEHLKQKSMT-----PAEREENQRMVQVQVWKYIILDKIDREKQAANKRKEES 2585
Db 2616 LSAVERLTQTSQGSTHFKEPASGDEGKIVL-----KORAKKKPNPVE 2658
Qy 2586 VEQKRSKQNA-----TKLSALLFKHKEOLRA-----EILKKRA 2618
Db 2659 EEPSSRRPRAPKAKAQPLEDLAGFTELSETSGHTQESLTAGKATKIPCESPPLEVDVTTA 2718
Qy 2619 LLDKDLQIEVQEEKRLDKIKKEDKDLMLAQ-----ATAV 2653
Db 2719 STKRHLRTRVOK-----VQVKEPSAVKFTQTSGETTDADKEPAGEDKGIKALKESAKQT 2773
Qy 2654 AAFCPPTVPLPAPPAPPPSP-----PPPGVQHTGLLSTPTPLPVASOKKRKEEKD 2705

Db 2774 PAPAASVTGSRRRPRAPRESAQAIJEDLAGFKDPAAGHTEESMTDDKTKTKIPCKSPELED 2833
Qy 2706 SSSKSKKKKMIKSTTSKTKKTKLYCICKTPYDESKFVIGCDRCQNNYHGRVCVGLQSEA 2765
Db 2834 TATSSKRRPRTRAKQVEVKEE--LLAVGK-----LTQTSG 2866
Qy 2766 ELIDYVYVCPQOSTEDAMTVLPTLTKDYEGKRLVLSLQAHKMAWPFLEPVDPN-DAPD 2824
Db 2867 E-----THTDKPEVGEKGTKA-----FKQAKRNVDAED 2897
Qy 2825 YYGVIKPEM-----DLATMEERVQR-RYYEKLTEFVAD 2856
Db 2898 VIGSRROPRAKPEKAQPLEDLASFQELSQTPTGHTTEELANGAAD 2940

RESULT 12
S37431
ankyrin 2, neuronal long splice form - human
N:Alternate names: ankyrin B, 440K splice form; ankyrin-B; brain ankyrin; non-erythro
N:Contains: ankyrin 2, short form
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence.Revision 06-Jan-1995 #text.Change 13-Aug-1999
C:Accession: S37431; A39643; B39643; A40334; A49462; S14533; S14569
R:Chan, W.
submitted to the EMBL Data Library, September 1993
A:Reference number: S37431
A:Accession: S37431
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3924 <CHA>
A:Cross-references: EMBL:Z26634; NID:9406287; PIDN:CAA81387.1; PID:9406288
R:Otto, E.; Kunimoto, M.; McLaughlin, T.; Bennett, V.
J. Cell Biol. 114, 241-253, 1991
A:Title: Isolation and characterization of cDNAs encoding human brain ankyrins reveal
A:Reference number: A39643; MUID:91302466; PMID:1830053
A:Accession: A39643
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2077 <OTL>
A:Cross-references: GB:X56957
A:Accession: B39643
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1443,3585-3924 <OTT>
A:Cross-references: EMBL:X56958
R:Tse, W.T.; Menninger, J.C.; Yang-Feng, T.L.; Francke, U.; Sahr, K.E.; Lux, S.E.; Wa
Genomics 10, 858-866, 1991
A:Title: Isolation and chromosomal localization of a novel nonerythroid ankyrin gene.
A:Reference number: A40334; MUID:92009921; PMID:1833308
A:Accession: A40334
A:Molecule type: DNA
A:Residues: 463-474,'PE',477-495 <TSE>
A:Cross-references: GB:X37123; NID:9178647; PIDN:AAA62828.1; PID:9178648
R:Chan, W.; Kordeli, E.; Bennett, V.
J. Cell Biol. 123, 1463-1473, 1993
A:Title: 440-KD ankyrinB: structure of the major developmentally regulated domain and
A:Reference number: A49462; MUID:94075409; PMID:8253844
A:Accession: A49462
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3924 <RES>
A:Cross-references: EMBL:Z26634; NID:9406287; PIDN:CAA81387.1; PID:9406288
C:Genetics:
A:Gene: GDB:ANK2
A:Cross-references: GDB:127607; OMIM:106410
A:Map position: 4q25-4q27
C:Superfamily: ankyrin; ankyrin repeat homology
C:Keywords: alternative splicing
F:2-3924/Product: ankyrin 2, long form #status predicted <MA>
F:2-1443,3585-3924/Product: ankyrin 2, short form #status predicted <MA2>
F:63-95/Domain: ankyrin repeat homology <AN01>
F:96-128/Domain: ankyrin repeat homology <AN02>

QY	754	GSVHGSKV---	LTSTLRLTLITQLENNIPSPFFHPNWSHRANIWKAQVCKSPREFALA	810
Db	711	MSDTGVSLSANLSEROLQVIN--SGDIP		739
QY	811	LAILECAVKPVMLPIWREFLGHRLHRMTSIEREKEVKK---	KEKKQBEETMQ---	864
Db	740	-----PITEILGEKVLSTRNAKQSDRYSASPTLRRSIRKHWNTQTP	785	
QY	865	-----QATWVKYTFPVK---	HQVMK--QKGEERYVGTGYGWSWISKTHVYRVFKLP	911
Db	786	KNVHNITDLEKTPVSETEPLKTASSVSKLRRSRLRHT-----	LVMETMNEKTBVILA	838
QY	912	GNTVNYRKSLEGT--KNMNDENMDESK--RACSRSPKIKIETP	OSEKDEVKSGDAAG	967
Db	839	ENTTARH---LRGTFREQVDOQODNENAPQRCEG---	ELSEGEKTSARRSSA---	889
QY	968	ADQNEWDISKITEKKDQDKVLLSDSDKPKCEPMEVDMDKMTESHVNCQE---	SSQVD	1024
Db	890	-----RKQKPTKDLGS-----	QWVTQADYAEELLSQGQT	921
QY	1025	VVNWSEGHRLTSYKKTSSKLDGLLERRIK--QFTLEK		1075
Db	922	IQNLEESMHM-----QNTSISEDQIGTEKKVNIIVYATKEKHSPKTPGKKAQPLEGPAGL	976	
QY	1076	KGIGTSTNSKNLSE-----SPVIT-----	KAKEGQCSMSRQECS---	1112
Db	977	KEHFPNPKDKPITEDTRVLCSPOVTTIINTTKPQTSISGKKVD--MKEESSALT	1035	
QY	1113	-----PNANNDOPEDL-----	IQGCSQSDSVL--RMSDPSHTTNKLYPKDRVLDVSR	1160
Db	1036	RHMPEGESRHPKILKECEDIKALKQSENEMLTSTVNGSKRTLGKSKKAQPLEDLT--	1093	
QY	1161	SPETCKPKO--NSTENDIEKVSDLASRGQEPKTS-----	KTGNDRPIDDSKLASAD	1211
Db	1094	-----CFQELFSPVPTNIIKKI-----	PKSPHTQPVRTIPASTKRLSKTGLSKVD	1139
QY	1212	DIGTLICKNKKPLQIEESDTIVSSSKS---	ALHSSVP--KSTNDRDA--TPLSRAMDFEG	1264
Db	1140	-----VROEPESTLGKRTKSPGRAGTPAPVQEBNDOCTAYMETPKOKLES	1186	
QY	1265	KLGCDSSENS--TLENSSDTVSIODSSEDMIVONSNE-----	SISEQFRTREQD	1312
Db	1187	LGRLKQKTPKIDITGFQDSQIIPDHANGPLVVVYTKMFNFSPQESAITRSRERQSR	1246	
QY	1313	VEVLE--PLKCELVSGE--STGNCEDRLPVKRGTEA--	NGKKPSQOKKLEERPYNKCSDOI	1366
Db	1247	ASISKIDVKEELLESEEHQLQEGVDTFQVSTNKNVIRSRKPAKR-----	1291	
QY	1367	KLKNTDKNNENRESEKKGORTSTFOINGKDNPKI--	YLKGECLKEISESRVVSNGVE	1424
Db	1292	KLDSTAGMNPNSKRMRCSSKNDTPCLEDLNGFQELFQMPGYANDSLTTGISTMLARSPQLG	1351	
QY	1425	PKVNNIN--KIIPENDIKSLTVKESAIRPFINGDIMEDFERNESSETPKSHLLSSSAEGN	1483	
Db	1352	PVYQTINKLSPKIILRKMDVTEE-----	ISG-----LWKQSLGRVHVHTQOEODN	1396
QY	1484	YRDSLETLPSTKE-----SDSTQTT-----	TPSASCPSNSVNVQVEDMEI--ETSEVKKVT	1532
Db	1397	AIKAIMEIP--KETLQTAADGTRLTROPQTP-----	KEYQPLEDSHSVFQELFQTSRYC	1448
QY	1533	SPSPITSEEBNSLS-----	NDFIDENGLPINKNNVNGE---	1565
Db	1449	SDPLGNKQTRMSLRSPQGVFVRPTRSKRLAKTSVGNIAVREKISPVSLPQCATGEVHV	1508	
QY	1566	-----SKRKTIVETVMTSTVATESKTVIKVEKDKO-----	TVVSSSTE	1605
Db	1509	IPIGEDDDTENKGV--KESTPQTLDSASRTVSKRQOGAHERPQFSGDLFHPQELFQTP	1566	
QY	1606	NCAKSTVTTTTTVTKLSTPSTGSDVLIISVKEQSKTVITVTVDLSLTITGGTLVTSMTV	1665	
Db	1567	ASGKDPTVDETTTKIALQSPQPGHIIINPASMROS-----	NMSL	1605
QY	1666	SREYSTRDVKLMKFSRPKTRSGTALPSYRKFTVTKTKSIFVLPNDLKLKARKGGIR	1725	

Db 2604 --SRRLRTHKGGWSTLLKLGDSKEITQ---ISDHSEKLAHDTSLIKTSQQQKP---DS 2655

Qy 2552 NORMIVGNQVMK-----YILDKID-----KEEQAAKKRK--REESVEQKRK-----2591

Db 2656 VKPLRTCRRLRASKEVPKEVLVDTRDHATLQSSNPLLSPKRKASRGDSIVRTRALRSL 2715

Qy 2592 --KONAYKLSALLPKFKHEQRAEILKKRALLDKDLQI-----EVQBELK-----RDLIK 2638

Db 2716 APKQEATDEKPVPEKKRAASSKRVSPPEVKMKHLKIVSNKLESVEEQVSTVMKTEBWA 2775

Qy 2639 KKERDLMQLAQATAVAACPVPVT-----VLPAPPAPPSPPPPPGVOHTGL-LSTTLP 2692

Db 2776 KREN-----PVTDPQNSRYKKTNNVKOPRPKFDAENYGIKKNEKTMK 2819

Qy 2693 VASQKRREBEKDSKSSKKKKMIS--TTSKETKKDTKLYCICKTPTYDESK 2741

Db 2820 TASQETELQNPDDGAKKSTSRGVSGKRTCLRSRGTTTETMPQPCAEKTSK 2870

RESULT 14

T34434

hypothetical protein K06A9.1a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000

C:Accession: T34434

R:Gelsel, C.; Gattung, S.

A:submitted to the EMBL Data Library, December 1996

A:Description: The sequence of C. elegans cosmid K06A9.

A:Reference number: Z1525

A:Accession: T34434

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2232 <GEI>

A:Cross-references: EMBL:U80846; PTDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a

A:Experimental source: strain Bristol N2; clone K06A9

C:Genetics:

A:Gene: CESP:K06A9.1a

A:Map position: x

A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049

Query Match 2.88; Score 419.5; DB 2; Length 2232;

Best Local Similarity 21.88; Pred. No. 6.6e-08;

Matches 339; Conservative 172; Mismatches 633; Indels 413; Gaps 63;

Qy 1071 LEGGIKIGIKGTSTNSSKNLSSESPIITRAKEGCOSDSMRQEQ---SPNANDQPEDLLIQG 1127

Db 731 LESSTSG-ATYSSGAGTTTSP-----SQSSSVGSSQGSTSPAASITSGEMTSQGS 781

Qy 1128 SQDSSSVLRMSDPSTHTNKLYPKDRVLDDVYSIRPETKCPQKNSIENDIEEKVSDLASRG 1187

Db 782 TQTPGSSVSTSAAILTSTQ-----QSVSTNSPGSTVTRPST-----VSGSTSSG 825

Qy 1188 QEPFKSKTKGNDFIDDSKLASADDIGTLCKNNKKPLQIESDTIV-----SSSKSALHS 1242

Db 826 STTVTVGSTAS---TSGSVASSSPAPS-TSQPNPNTSSGSSMITQSPYSPQSTSPVES 881

Qy 1243 SVPKSTNDRATPLSRAMDFEGLCGDSENSNTLENS---SDTVSIQDSSEEDMIVQNSN 1299

Db 882 STTPSPGSGPTTLTSTS-----PSPQSTTIGTQGSTSPGISTTSEE-MTSQGST 931

Qy 1300 ESISEQFRTREQDVEVLEPLKCELVGSESTGNCEDRLPVKGTANGKKPQQKKLERPV 1359

Db 932 QTPGSGTGTVPQSTVSD-----STSSGS---TTVTGVSTEGS-SSPIPSTQNTNPS 979

Qy 1360 NKCSQDLKLTNTDKNKNNRESEKKQORTSTFQINGKDNKPKIYLKGECLKSEISERV 1419

Db 980 TSSGSSMSTQTPQSSQSTSPVESSTGATSSS-----GSP-----GTTLTSTSPSPSP 1027

Qy 1420 S-----GNVEPKVNNINKIIE-NDIKSLTVKESAIRPFINGDVIMEDFNERNSET 1470

Db 1028 SSTIGSQSGSTSPVSTVTSQGSTGTGPGSTGTVPKPT-----VSGSSAGSGSTATMGSTEA 1083

Qy 1471 KSHLLSSSDAEGNVRDSLETLPTGK-----ESDSTQTTTPSACQPSNNOVEDWEIETSE 1527

[illegible]

Db 2011 TMSSGATTSGDKMFLSTGTTTYSFSSRGSLATTSAAPKPSVTCLFMYDTQSKEIDQ 2067

RESULT 15

D71623

erythrocyte membrane protein pFEMP3 PFB0095c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: D71623

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Perle, M.; Salzberg, S.; Zhou, L.; Sutton, G.C.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: D71623

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2441 <GAR>

A:Cross-references: GB:AE001371; GB:AE001362; NID:g3845092; PIDN:AAC71809.1; PID:g384509

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0095c

Query Match 2.8%; Score 417; DB 2; Length 2441;
Best Local Similarity 18.6%; Pred. No. 9.1e-08;
Matches 525; Conservative 424; Mismatches 1053; Indels 824; Gaps 132;

Qy 215 YGPEVKIKVQLVDFQTLNAREELMSEG-----IQYDDHCRVCHKLGLD 263

Db 44 YNKIDNYNIFELRLKSLAQVLGNLRSRGVRDPTKEALKEKFRDHRK----- 95

Qy 264 LCCTCSAVYHLECVPLPEVEPEDEWQCEVCV-----AHKVPGVDCVAEIQ 311

Db 96 -----KEALKQKTEKNEKARNALKKEKQKNDQAQAKDLTKESQDS 140

Qy 312 KNPYIRHEPIGVDRSRKWFNRLRIIEEDFENENEKIWIYSYKVG----- 360

Db 141 SSEKSLKEVNGEALKE-----NKTLLKKELENOKEB-----EKNKIKDNDEALKNG 193

Qy 361 -----LAELIDCLDKYAEALCKILEEMREEIHRHMDITEDLTNKAAGNSFLAA 412

Db 194 NDKDDKIVPKPESVEKDLKEMEL-----KEFEIKQHLKDYEERKEKRR----- 239

Qy 413 ANEILSIRAKGIDINVKSPETEKNKETE-----NSDKDAEKNEEF-----EDQSLEKDS 467

Db 240 --NWILRSLR-----DKLEIEQLKNAQLESAINELKERRASRRPMVMKMQRMKDE 292

Qy 468 DD---KTPDDDPQOGKSEEPTEVGDGKNS-----VSANIGDNTTNATSEETSPSEGRSPVG 520

Db 293 VDEWIKYDDEQAENKTKDDEEIKDGDGYEEIVETKFGYGMRENA----- 337

Qy 521 CLSETPDSSNMAEKK-----VASELPQDVPEEPNKTCESSNTSATTTTSIOPNLSENS 573

Db 338 -LGELDEVEERYEKKRYLLKEDGBGL- KDVEEKLEETGYGFREKPTTILYKRRNKE 395

Qy 574 SSELNSQSSEAKAADPENGERSHTPPVSIQIEIVGDFTSEKTSBELSPGAGGASG 633

Db 396 QKKLKEDKEKLLIAAEBPDD- EKIKLKDS- DOKVVVVPVNNKSS----- 438

Qy 634 STRIITLRPDSKLSQKLSQVAAAHAANKLF-----KEGKEVLVYNSQGEISRLSTKK 689

Db 439 ---FPDFRAPDKKRTMF-----YRLSELFPVPRKDNE-LAVSGDCMSKVNGKK 485

Qy 690 EVIMKGINNYFKLGQEGKYRVYHNOYSTNSFALN--KHQHRHDKRRHLAHKFCULTPA 747

Db 486 ---LKSTFNPFRNRNKLKER---KMQLHKFKKNYKQKLLEREREN-----PD 531

Qy 748 GEFKNGSVHGSKVLTTISTLRLLITQLENNIPSFPHNPWASHRAWIKAVQMCSPREF 807

Db 532 GEPLNTPEIH---VIREPSDL-----MDKGNSAGHP-FKYQPTKGLKEYEESHVAKDY 581

Qy 808 ALALAILECAVKPVVMLP INREFLGHTRLRMTSIEEREKEKVKKEKQBEETMQOAT 867

Db 582 QL-----EHEPPTKLPYEYK--CH--VSREYQLDNVRODELPEYKCHVSRE----- 624

Qy 868 WVKYTFPVKQKQKGEYRVTYGGWSWISKTHVYRVFPKLPNGTNNVYRKSLGKTK- 926

Db 625 -----YQLDNEV-RDELPEY-----EKGHVSR-----EYQLDNEGPST 656

Qy 927 -----NNMDENMDSRKRKSRSPKIKIIPDSKDEKDEKGSDAAGK--- 967

Db 657 LKEYDQTELAKGKIDITNKPHEVSDYDQSELAKG-KDITNKPHEVSDYDQTELAKREV 715

Qy 968 -----ADQEMDISK---LITEKKDQVKELLDSD--SDKPCKEPEWE-VDDDMKTE- 1012

Db 716 TNKPHENLEEYNETDLAKGEVNTNKPHEVSDYDQSELAKGKIDITNKPHEVSDYDQTEL 775

Qy 1013 -----SHVNCQESSQVDVYVNSEGFHLRTSYKKTKSKSLDGLLER-----RIK 1056

Db 776 AKGKEVTNKPHEYLEYNETDLA-----KGKEVTNKAHENLEEYNETDLAKGK 823

Qy 1057 QTTLEEKORLEKIKLEGGIKIGKSTNSS-KNLSE-----SPVITRAKEGC----- 1102

Db 824 EYTNKAHENLEEYNETDLAK--GKEVTNKAHENLEEYNETDLAKGEVNTNKAHENLEEY 881

Qy 1103 QSDSMRQEQSPNANDQPE-----DLIOGCSQSDSVLRMSDPST-----TNKLYP 1149

Db 882 ETDLAGKEVNTNKAHENLEEYNETDLAKGEVNTNKAHENLEEYNETDLAKGEVNTNKAHE 941

Qy 1150 KDRVLDDVSIRSPETKCPKQNSIENDIEEKV-----SDLASRGQPTKSKTKGNDFIDD 1204

Db 942 N--LEEYN-----ETDLAKGEVNTNKAHENLEEYNETDLA-KGKEVT-NKARENLEEY 992

Qy 1205 SKLASADDTGLTICKKPKLIQEESDTIVSSSKSALHSSVPKSTNDROATPLSRAM---- 1260

Db 993 TDLAKGEVNTNKAHENLEEYNETDLAKGEVNTNKA--RENLEEYNETDLAKGEVNTN 1047

Qy 1261 -----DPEGLKGDGSDSNTLENSDTSVTDSSSE-----EDMIVQNSNESIS 1303

Db 1048 KARENLEEYEEK--DYMKNNELQNGSD-GLKENAELKNKELNRKNGSDGLKENA-ELKN 1102

Qy 1304 EGFRTREQD--VEVLEPLKCELVSESTG---NCE---DRLPVKTEANGKKPSOOKLE 1355

Db 1103 KELNRKNGSDGLKENAELKNKELQNGSEGLKENAELKNKELQNGKSE--GLKENAELKNK 1160

Qy 1356 ERPVNCSDQIK-----LKNNTDKKNE-----NRBS 1382

Db 1161 ELR-NKSGEGLKENAELKNKELQNGSEGLKENAELKNKELQNGSEGLKENAELKNKEL 1219

Qy 1383 EKKGO-----RTSTFQINGKON-KPKIYKLGECLEKEISERVVSGNVPKVNINK 1432

Db 1220 QNGSGEGLKENAELKNKELNRKNGSEGLKENVYTNNDLKNNDIQNKDLSNKMKNKELLNK 1279

Qy 1433 IIPENDIKS-----LTVKESAIRPFINGDVIMEDFNERNSETKSHLLSSDAEGNYR 1485

Db 1280 DISNKMKNKELNKLNDKSNEDMKNELNKLNRKDLKSGNMEQONTGLKNTPSKGOON 1339

Qy 1486 DSIETLPTKESDSTQTTTPSASCPSNSVNOVEDMEIETSEVKKVTSSPITSEBSNLS 1545

Db 1340 TGLKNTPNEROQNTGLKNTPS-----EQQNTGLKNTPEQOQNTGLKNTPNEROQNTGLK 1395

Qy 1546 N---DFIDENGLPINKNENVNGESKRKTIVITEVTMTSTVATESKTVIIVKRGDKQTVVS 1602

Db 1396 NTPSEGOQNTGL--KNTPIEGQON-----TGLKNTPS-----EQQNTGLK 1434

Qy 1603 STEN-----CAKSTVTTTTVTKLSTPSTGSGVDIISVKBSKTVVTTVTDLSLTNG 1656

Db 1435 NTPNEROQNTGLKNAANKGOQNTGLKNTPSKG-----QONTGLKNTPNEROQNTG 1484

Qy 1657 GTLVTSMTVSKEYSTRDKVKLMKFRPKKTRSGTAL---PSYRKFEVTKTKKSIFVLND 1713

Db 1485 ---LKNTPNEROQNTGLK-----NTPSEGOQNTGLKNTPNERQ-----QNT 1522

Qy 1714 DLKKLARKG---GIREVPYFNNAKPAIDWPSPRPTFGITWRYLQTVKSLAGVSL 1769

Db 1523 GLKNASKGOQNTGLKNAP----- 1541

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2003, 16:10:05 ; Search time 53.7816 Seconds
(without alignments)
8207.607 Million cell updates/sec

Title: US-09-698-295-10

Perfect score: 14333

Sequence: 1 MYSEEEEDGDAEETQDSE.....KLKFKASRSHNNKLQSTAS 2781

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14333	100.0	2781	21 AAY57453	Human transcriptio
2	14170	98.9	2907	21 AAY57452	Human transcriptio
3	13428	93.7	2759	24 AAO16418	Human nucleic acid
4	2695.5	18.8	557	22 AAB94078	Human protein sequ
5	1557.5	10.9	976	22 AAB68735	Drosophila melanog
6	1474	10.3	328	21 AAB54364	Human pancreatic c
7	1030	7.2	191	24 ABU70846	Human adipocyte se
8	982.5	6.9	238	20 AAY04323	Human secreted pro
9	981	6.8	237	22 AAU19446	Human diagnostic a

10	789	5.5	149	23 AAU78845	SCAN/KRAB protein
11	781	5.4	246	22 AAU31944	Novel human secret
12	672.5	4.7	997	22 ABB62887	Drosophila melanog
13	572.5	4.0	2768	22 ABB68397	Drosophila melanog
14	560	3.9	5533	22 ABB65772	Drosophila melanog
15	560	3.9	5560	22 ABB71160	Drosophila melanog
16	549.5	3.8	6815	22 ABB66811	Drosophila melanog
17	534	3.7	3080	22 ABB64877	Drosophila melanog
18	529.5	3.7	3257	22 ABB67502	Drosophila melanog
19	508.5	3.5	2897	22 ABB58514	Drosophila melanog
20	441.5	3.1	4498	22 ABB58595	Drosophila melanog
21	437	3.0	3021	24 ABB75873	Human secretory po
22	434.5	3.0	2977	22 ABB69480	Drosophila melanog
23	432.5	3.0	1878	22 AAM40239	Human polypeptide
24	432.5	3.0	2137	23 ABB39618	Staphylococcus epi
25	432.5	3.0	3266	21 ABA2491	Human ORFX ORF2255
26	430.5	3.0	3684	24 ABR47592	Breast cancer asso
27	430	3.0	1795	22 ABB69806	Drosophila melanog
28	430	3.0	2586	22 ABB66878	Drosophila melanog
29	429	3.0	3726	22 ABB63947	Drosophila melanog
30	428	3.0	1727	22 AAB95554	Human protein sequ
31	423	3.0	2285	22 ABB63057	Drosophila melanog
32	418	2.9	4019	22 ABE13839	Human lung tumour-
33	418	2.9	4025	23 ABB69736	Human polypeptide
34	416.5	2.9	2665	22 ABB48336	Human liver peptid
35	416.5	2.9	2665	22 ABB28314	Human peptide #965
36	416.5	2.9	2665	22 ABB33490	Peptide #996 encod
37	416.5	2.9	2665	22 ABB18950	Protein #949 encod
38	416.5	2.9	2665	22 AAM54270	Human brain expres
39	416.5	2.9	2665	22 AAM66665	Human bone marrow
40	416.5	2.9	2665	22 AAM14533	Peptide #967 encod
41	416.5	2.9	2665	22 AAM26950	Peptide #987 encod
42	416.5	2.9	2665	22 AAM02259	Peptide #941 encod
43	416.5	2.9	2665	23 ABB36319	Human peptide enco
44	416	2.9	10431	24 ABU54861	Human CAL25 amino
45	415	2.9	3236	24 ABB96155	Human Ki-67 chromo

ALIGNMENTS

RESULT 1
AAY57453
ID AAY57453 standard; Protein; 2781 AA.
XX
AC AAY57453;
XX
DT 22-FEB-2000 (first entry)
XX
Human transcriptional regulatory factor SEQ ID NO:10.
XX
DE
XX
Human; transcriptional regulatory factor; TCoal; BLAST detection;
KW bromo-domain; cell proliferation; cancer.
XX
XX Homo sapiens.
OS
XX
WO9957143-A1.
PN
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-JP02340.
XX
PR 30-APR-1998; 98JP-0137631.
XX
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
XX Jones MH;
XX
XX WPI; 2000-052940/04.
XX
XX N-PSDB; AA239033.
XX
XX Transcriptional regulatory factor containing a bromo domain and gene
PT TCoal encoding it

xx Claim 1; Page 139-151; 154pp; Japanese.

xx The present sequence represents a human transcriptional regulatory factor

xx containing a bromo domain. The factor interacts with proteins involved

xx in the chromatin-mediated transcription regulatory mechanism. It binds

xx to hSNF2H, hSNF2L and NcoA-62/Skip. It can be used for screening

xx compounds binding to it and acting as agonists or antagonists, which

xx are potentially useful for the treatment and prevention of cancer and

xx other cell proliferation disorders.

xx Sequence 2781 AA;

Query Match 100.0%; Score 14333; DB 21; Length 2781;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2781; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVSSEEEEDGDAEETQDSEDEDEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHST 60

Db 1 MVSSEEEEDGDAEETQDSEDEDEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHST 60

Qy 61 YSSTPGRRKPRVHRPRSPILAEKDIPLLEPKSSSEDLMPNEHIMNVIAIYEVLRNFGTV 120

Db 61 YSSTPGRRKPRVHRPRSPILAEKDIPLLEPKSSSEDLMPNEHIMNVIAIYEVLRNFGTV 120

Qy 121 LRLSPFRFEDFCAALVSOEQCTLMAEMHVLLKAVLREEDTSNTTGFADLKDSVNSTLY 180

Db 121 LRLSPFRFEDFCAALVSOEQCTLMAEMHVLLKAVLREEDTSNTTGFADLKDSVNSTLY 180

Qy 181 FIDGWTWPEVLRYCESDKYHHVLPYQEAEDYYPGVENKIKVLQFLVQFLTNTARE 240

Db 181 FIDGWTWPEVLRYCESDKYHHVLPYQEAEDYYPGVENKIKVLQFLVQFLTNTARE 240

Qy 241 ELMSEGVIOYDHCRCVKHGLDLCCTCSAVYHLECVKPLEEVPEDMOCVCAHV 300

Db 241 ELMSEGVIOYDHCRCVKHGLDLCCTCSAVYHLECVKPLEEVPEDMOCVCAHV 300

Qy 301 PGVTDCAEIQKPKYIRHEP IGYDRSRKRYWFLNRRLIIBEDTENENKKIWTYSTKVQ 360

Db 301 PGVTDCAEIQKPKYIRHEP IGYDRSRKRYWFLNRRLIIBEDTENENKKIWTYSTKVQ 360

Qy 361 LAELIDCLDKDYWEALCKIILEMREETHRMDITEDLTNKARSNKSFLLAAANEELLES 420

Db 361 LAELIDCLDKDYWEALCKIILEMREETHRMDITEDLTNKARSNKSFLLAAANEELLES 420

Qy 421 IRAKKGIDINVKSPETEKKNETENDSKAEKNREFEFQSLKSDDKTPDDDPQGGK 480

Db 421 IRAKKGIDINVKSPETEKKNETENDSKAEKNREFEFQSLKSDDKTPDDDPQGGK 480

Qy 481 SEVGDFKSEKNGELSESPGAGKASGSTRIITRLNPDSKLSQLKSOQVAAAAHEANKL 540

Db 481 SEVGDFKSEKNGELSESPGAGKASGSTRIITRLNPDSKLSQLKSOQVAAAAHEANKL 540

Qy 541 FKEGKVLVNSQGEISRLSTKKEVIMKGNINNYFKLQEGKYRVYHNOYSTNSFALNKH 600

Db 541 FKEGKVLVNSQGEISRLSTKKEVIMKGNINNYFKLQEGKYRVYHNOYSTNSFALNKH 600

Qy 601 QHREDHDKRRHLAKHFKCLTPAGEFKWNGSVHGSKVLTITLRLITQLENNIPSSFLHPN 660

Db 601 QHREDHDKRRHLAKHFKCLTPAGEFKWNGSVHGSKVLTITLRLITQLENNIPSSFLHPN 660

Qy 661 WASHRANNIKAVQMSCKPREPALAIAILECAVKPVMLPIWREFLGHTRLHRMYSIEREE 720

Db 661 WASHRANNIKAVQMSCKPREPALAIAILECAVKPVMLPIWREFLGHTRLHRMYSIEREE 720

Qy 721 KEVKKKKKKEEETMOQATWVKYTFPVKHQVWKQGEERYVTCYGGWSWISKTHVYRF 780

Db 721 KEVKKKKKKEEETMOQATWVKYTFPVKHQVWKQGEERYVTCYGGWSWISKTHVYRF 780

Qy 781 VPKLPNGTNVNYRKSLGCTKNMNDENMDESCKRCSRSPKKIKIEPDSEKDEVKGSAAK 840

Db 781 VPKLPNGTNVNYRKSLGCTKNMNDENMDESCKRCSRSPKKIKIEPDSEKDEVKGSAAK 840

Qy 841 GADQNMWDISKITEKKDQDYKELLDSDSKPKKEEPMVEYDDDMKTESHVNCQESSQVDVV 900

Db 841 GADQNMWDISKITEKKDQDYKELLDSDSKPKKEEPMVEYDDDMKTESHVNCQESSQVDVV 900

Qy 901 NVSEGFHLRTSYKKTKSSKLDGLLEPRRIKQFTLEEKQORLEKIKLEGGIKGIGKTSNSS 960

Db 901 NVSEGFHLRTSYKKTKSSKLDGLLEPRRIKQFTLEEKQORLEKIKLEGGIKGIGKTSNSS 960

Qy 961 KNLSESPVITKAKEGCQSDSMRQEQSPNANNDQPEDLIQGCQSDSDSVLRMSDPSHTNK 1020

Db 961 KNLSESPVITKAKEGCQSDSMRQEQSPNANNDQPEDLIQGCQSDSDSVLRMSDPSHTNK 1020

Qy 1021 LYPKDRVLDDVYSRSPETKCPKQNSIENDIEEKVSDIASRQBPETKSKTKGNDDFFIDDSK 1080

Db 1021 LYPKDRVLDDVYSRSPETKCPKQNSIENDIEEKVSDIASRQBPETKSKTKGNDDFFIDDSK 1080

Qy 1081 LASADDITGLICKNKPKLIQEEEDTIVSSSKSALHSSVPKSTNDROATPLSRAMDFEGLK 1140

Db 1081 LASADDITGLICKNKPKLIQEEEDTIVSSSKSALHSSVPKSTNDROATPLSRAMDFEGLK 1140

Qy 1141 GCDSESNSTLENSDVTYSIQDSSSEEDMIVQNSNESISEQFRTREQDVEVLEPLKCELVS 1200

Db 1141 GCDSESNSTLENSDVTYSIQDSSSEEDMIVQNSNESISEQFRTREQDVEVLEPLKCELVS 1200

Qy 1201 ESTGNCEDRLPVKGTTEANGKKPSQOKKLEERPYNKCSDOIKLKNTTDDKKNNENRESEKK 1260

Db 1201 ESTGNCEDRLPVKGTTEANGKKPSQOKKLEERPYNKCSDOIKLKNTTDDKKNNENRESEKK 1260

Qy 1261 QRTSTFOINGKDNKPKIYLKGECLKEISESRVSGNVPEKPVNNINKIIPENDIKSLTVKE 1320

Db 1261 QRTSTFOINGKDNKPKIYLKGECLKEISESRVSGNVPEKPVNNINKIIPENDIKSLTVKE 1320

Qy 1321 SAIRPFINGDVIMEDFNERNSSETKSHLLSSDAEGNYRDSLETLPSTKESDSTQTTPS 1380

Db 1321 SAIRPFINGDVIMEDFNERNSSETKSHLLSSDAEGNYRDSLETLPSTKESDSTQTTPS 1380

Qy 1381 ASCPESNSVQVQVEDMEIETSEVKVTSPTISEESNLNDFIDENGLPINKNENVAGES 1440

Db 1381 ASCPESNSVQVQVEDMEIETSEVKVTSPTISEESNLNDFIDENGLPINKNENVAGES 1440

Qy 1441 KRKTIVIEVTMTSTVATESKTVIKVEKGQKQTVVSTENCACAKSTVTTTTVTTKLSTPS 1500

Db 1441 KRKTIVIEVTMTSTVATESKTVIKVEKGQKQTVVSTENCACAKSTVTTTTVTTKLSTPS 1500

Qy 1501 TGSVDIISVKEQSKTVVTTTDTLTTGGTLVTSMTVSKYSTROKVKLMKFSRPKT 1560

Db 1501 TGSVDIISVKEQSKTVVTTTDTLTTGGTLVTSMTVSKYSTROKVKLMKFSRPKT 1560

Qy 1561 RSGTALPSYRKFTVKTKSIFVLPNDDLKKLARKGGIREVPYPYNAKPALDIWYPSP 1620

Db 1561 RSGTALPSYRKFTVKTKSIFVLPNDDLKKLARKGGIREVPYPYNAKPALDIWYPSP 1620

Qy 1621 RPTFGITWRYRLQTVKSLAGVSLMLRLIMASLRWDDMAAKYPPGGGSTRTETSETEITT 1680

Db 1621 RPTFGITWRYRLQTVKSLAGVSLMLRLIMASLRWDDMAAKYPPGGGSTRTETSETEITT 1680

Qy 1681 EIIKRRDVGPIRFEYCIRKIIICPIGVPEPKETPTPQKGLRSSALRPKRPETPKQT 1740

Db 1681 EIIKRRDVGPIRFEYCIRKIIICPIGVPEPKETPTPQKGLRSSALRPKRPETPKQT 1740

Qy 1741 PVIIETWVAEELELEWEIRAFAEVEKEKAQAVQQAQKRLQEQKPVVIATSTSPSSST 1800

Db 1741 PVIIETWVAEELELEWEIRAFAEVEKEKAQAVQQAQKRLQEQKPVVIATSTSPSSST 1800

Qy 1801 TSTISPAQKVMVAPISGVTGTGKWLTTTKVGSFATVTFQONKHFQTFATVWKQOGSNS 1860

Db 1801 TSTISPAQKVMVAPISGVTGTGKWLTTTKVGSFATVTFQONKHFQTFATVWKQOGSNS 1860

Qy 1861 GWVQVQKVLGIIPSSSTGTSQOTFTSQPRTATVTIIRPNTSGSGGTTNSQVITGPQIRP 1920

Db 1861 GWVQVQKVLGIIPSSSTGTSQOTFTSQPRTATVTIIRPNTSGSGGTTNSQVITGPQIRP 1920

Qy 1921 GMTVIRTPLOQSTGLGKAIIRTPVMVPGAPQVMTQIIRGPQVSTAVSPNTVSTSPQK 1980

Db	541	QDVPEPNKTCSSNTSATTTGIQWLENSNSSSELNSQSASAKAADDPENGRESHT	600
Qy	483	-----VGDFKSEKNGELSESPGAGKGASGSTRITRLRNPDSKLSQKQVAAAA	534
Db	601	PVSIQEEIVGDTSEKSTGELSESPGAGKGASGSTRITRLRNPDSKLSQKQVAAAA	660
Qy	535	HEANKLFKEGKVLVWNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKYVYINQYSTNS	594
Db	661	HEANKLFKEGKVLVWNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKYVYINQYSTNS	720
Qy	595	FALNKHQHRDHDRRLAHKFCLTPAGEFKWNGSVHGSKVLITSLRLTITOLENNIPS	654
Db	721	FALNKHQHRDHDRRLAHKFCLTPAGEFKWNGSVHGSKVLITSLRLTITOLENNIPS	780
Qy	655	SFLHPNASHRANWIKAVQCMCKPREPALALALECAVKPVWMLPIWREFLGHTLRHMT	714
Db	781	SFFHPNASHRANWIKAVQCMCKPREPALALALECAVKPVWMLPIWREFLGHTLRHMT	840
Qy	715	SIEREEKYKKEKKQEEETMOATWVYTPPVKHQVWKQGEERYVTGYGWSWISK	774
Db	841	SIEREEKYKKEKKQEEETMOATWVYTPPVKHQVWKQGEERYVTGYGWSWISK	900
Qy	775	THVRFVPKLPNGTNVNYRKSLEGTKNMNDENMDESCKRSRPPKIKIEPDSEKDEVK	834
Db	901	THVRFVPKLPNGTNVNYRKSLEGTKNMNDENMDESCKRSRPPKIKIEPDSEKDEVK	960
Qy	835	GSDAAKAGQONEMDISKITEKKQDVVELLDSDSKPCKEPEMVEDDDMKTESHVNCQES	894
Db	961	GSDAAKAGQONEMDISKITEKKQDVVELLDSDSKPCKEPEMVEDDDMKTESHVNCQES	1020
Qy	895	SOQDVVNVSEGFHLRTSYKKTYSKLDGLLERIKOFTLEEKORLEKIKLEGKIGIKG	954
Db	1021	SOQDVVNVSEGFHLRTSYKKTYSKLDGLLERIKOFTLEEKORLEKIKLEGKIGIKG	1080
Qy	955	TSTNSSKNLSEPVITKAKEGCQSDSMRQEQSPNANDQPEDLLQGCQSDSSVLRMSDP	1014
Db	1081	TSTNSSKNLSEPVITKAKEGCQSDSMRQEQSPNANDQPEDLLQGCQSDSSVLRMSDP	1140
Qy	1015	SHITNKLVPKRVLDVDSIRSPETPKPONSIENTIEEKVSDLASRGOEPTKSTKGNDF	1074
Db	1141	SHITNKLVPKRVLDVDSIRSPETPKPONSIENTIEEKVSDLASRGOEPTKSTKGNDF	1200
Qy	1075	FIDDSKLASADDIGTLCKNKKPLIQEESDTIVSSKSKALSHSVKPNSTNDRDATPLSRAM	1134
Db	1201	FIDDSKLASADDIGTLCKNKKPLIQEESDTIVSSKSKALSHSVKPNSTNDRDATPLSRAM	1260
Qy	1135	DPEGKLGCDSESNSTLENSDTVSIQDSSEEDMIVQNSNESISSEQFRTREODVEVLEPLK	1194
Db	1261	DPEGKLGCDSESNSTLENSDTVSIQDSSEEDMIVQNSNESISSEQFRTREODVEVLEPLK	1320
Qy	1195	CELVSSESTGNCEDRLPVKGTGANGKPSQOKKLEERPVPNKCSDQIKLKNNTDKKNENR	1254
Db	1321	CELVSSESTGNCEDRLPVKGTGANGKPSQOKKLEERPVPNKCSDQIKLKNNTDKKNENR	1380
Qy	1255	ESEKKGQRTSTFQINGKONKPKIYLGECLEKIESESRRVSGNVEPKVNNIKIIPENDIK	1314
Db	1381	ESEKKGQRTSTFQINGKONKPKIYLGECLEKIESESRRVSGNVEPKVNNIKIIPENDIK	1440
Qy	1315	SLTVKESATRPINGDVIMEDFNERNSETSKSHLLSSDAGNRYDSLETLPSTKESDST	1374
Db	1441	SLTVKESATRPINGDVIMEDFNERNSETSKSHLLSSDAGNRYDSLETLPSTKESDST	1500
Qy	1375	QTTTPSASCPESNVQVEDMEIETSEVKVYTSPTITSEESNLSNDFIDENGLPIKNKE	1434
Db	1501	QTTTPSASCPESNVQVEDMEIETSEVKVYTSPTITSEESNLSNDFIDENGLPIKNKE	1560
Qy	1435	NYNGESKRTVITEVTTWTSTVATESKVIKVEKGDKQTVVSSSTENCAKSTVTTTTVT	1494
Db	1561	NYNGESKRTVITEVTTWTSTVATESKVIKVEKGDKQTVVSSSTENCAKSTVTTTTVT	1620
Qy	1495	KLSTPSTGSGVDIIISVKEQSKTVVTTTSDSLTTTGGTLVTSMTYSKEYSTRDKVKLMKF	1554
Db			
Db	1621	KLSTPSTGSGVDIIISVKEQSKTVVTTTSDSLTTTGGTLVTSMTYSKEYSTRDKVKLMKF	1680
Qy	1555	SRPKTRSTALTSPSYRKFFVTSTKKSIFVLPNDDLKKLARKGIREVYFVFNNAKPALDI	1614
Db	1681	SRPKTRSTALTSPSYRKFFVTSTKKSIFVLPNDDLKKLARKGIREVYFVFNNAKPALDI	1740
Qy	1615	WYPSPRPFTGTTWRYRLQTVKSLAGVSLMLRLMLWASLWDDMAAKVPPGGSGSTTETSE	1674
Db	1741	WYPSPRPFTGTTWRYRLQTVKSLAGVSLMLRLMLWASLWDDMAAKVPPGGSGSTTETSE	1800
Qy	1675	TEITTTTEIIRKRDVGPYGRFECYRIRKIIICPIGVETPKETPTPORKGLRSALARKPE	1734
Db	1801	TEITTTTEIIRKRDVGPYGRFECYRIRKIIICPIGVETPKETPTPORKGLRSALARKPE	1860
Qy	1735	TPKOTGPVITETWVAEELELEWEIRAFARVEKEKAQAVEQOQAKRLEQOKPTVIATSTT	1794
Db	1861	TPKOTGPVITETWVAEELELEWEIRAFARVEKEKAQAVEQOQAKRLEQOKPTVIATSTT	1920
Qy	1795	SPTSSTTSTISPAQKVMVAPISGSVTTGTKMVLTKVGSPIVTFQONKNFHQTATWYK	1854
Db	1921	SPTSSTTSTISPAQKVMVAPISGSVTTGTKMVLTKVGSPIVTFQONKNFHQTATWYK	1980
Qy	1855	QGSNSGVVQVQKVLGIIIPSSGTSGOQTFSPQRTATVTIRPNTSGSGGTTNSOVIT	1914
Db	1981	QGSNSGVVQVQKVLGIIIPSSGTSGOQTFSPQRTATVTIRPNTSGSGGTTNSOVIT	2040
Qy	1915	GPOIRPGMTVIRTPLOQSTLGAIIIRTPVMVOPGAQVQVMTQIIRGQPVSTAVSAPNTVS	1974
Db	2041	GPOIRPGMTVIRTPLOQSTLGAIIIRTPVMVOPGAQVQVMTQIIRGQPVSTAVSAPNTVS	2100
Qy	1975	STPGQSKLSATSSTNSIQSSASQPPRPOQGVKLTMAOLTOLTQGHGGNGLTVLQOGG	2034
Db	2101	STPGQSKLSATSSTNSIQSSASQPPRPOQGVKLTMAOLTOLTQGHGGNGLTVLQOGG	2160
Qy	2035	QTTGQLQLIPQGVTVLPGPGQOQLMAAMPNGTVQREFLTPLATTATATSTTTTSTTAA	2094
Db	2161	QTTGQLQLIPQGVTVLPGPGQOQLMAAMPNGTVQREFLTPLATTATATSTTTTSTTAA	2220
Qy	2095	GTGEORQSKLSPQMVHQDKTLPPAOSSSVGPBAPAKAPQTAQPSARPOQTPQPSAPQEV	2154
Db	2221	GTGEORQSKLSPQMVHQDKTLPPAOSSSVGPBAPAKAPQTAQPSARPOQTPQPSAPQEV	2280
Qy	2155	QTOPEVQTOQTVSSHVPSEAPTHAOSKSKPQVAAQSQPOSNVQSQSPVRVQSPQTRIRP	2214
Db	2281	QTOPEVQTOQTVSSHVPSEAPTHAOSKSKPQVAAQSQPOSNVQSQSPVRVQSPQTRIRP	2340
Qy	2215	STPSQLSPGQOSQVQTTTSQPIPIQHTSLQIPSQGQPSQPOVQSTQTLSSGQTLNQV	2274
Db	2341	STPSQLSPGQOSQVQTTTSQPIPIQHTSLQIPSQGQPSQPOVQSTQTLSSGQTLNQV	2400
Qy	2275	SVSSPSRPOLOIQPOPOQVITAVPOLQOQOVVLSQIOSOVVAQIOAOSGVPQOIKLQLP	2334
Db	2401	SVSSPSRPOLOIQPOPOQVITAVPOLQOQOVVLSQIOSOVVAQIOAOSGVPQOIKLQLP	2460
Qy	2335	QIOQSSAVQTHQIONVYTVQAASVQEQLOQVQQLRDQOQKQKQOQIEINVTSPKLLIK	2394
Db	2461	QIOQSSAVQTHQIONVYTVQAASVQEQLOQVQQLRDQOQKQKQOQIEINVTSPKLLIK	2520
Qy	2395	EIIQKQVVMKHNAVIEHLKOKKSMTPAEREENQRMVNCQVMKYLDKIDKEKQOAKKR	2454
Db	2521	EIIQKQVVMKHNAVIEHLKOKKSMTPAEREENQRMVNCQVMKYLDKIDKEKQOAKKR	2580
Qy	2455	KREESVEQKSKONATKLSALLFKHKBOLRAELIKKRALDKDLOIEVQBELKRDLIK	2514
Db	2581	KREESVEQKSKONATKLSALLFKHKBOLRAELIKKRALDKDLOIEVQBELKRDLIK	2640
Qy	2515	EKDLMLQAQATAVAACPVPVTPVLPAPAPPSPPPPGVQHTGLLSTPLPVASQKRKR	2574
Db	2641	EKDLMLQAQATAVAACPVPVTPVLPAPAPPSPPPPGVQHTGLLSTPLPVASQKRKR	2700
Qy	2575	BEEDSSSKSKKKKMISTTSKETKDKLYCICKTPYDESKFYIGCDRCQNWYHGRCVGI	2634
Db	2701	BEEDSSSKSKKKKMISTTSKETKDKLYCICKTPYDESKFYIGCDRCQNWYHGRCVGI	2760

QY 2635 LOSEALIDEYVCPQCQSTEDAMTVLPTTEKDYGLKRVLSLQAHKMAWFFLEVPDNP 2694
|||||
DB 2761 LOSEALIDEYVCPQCQSTEDAMTVLPTTEKDYGLKRVLSLQAHKMAWFFLEVPDNP 2820
|||||
QY 2695 DAPDYGVYKPEMDLATMEERVORRYEYKLTTEFVADMTKIFDNCRYNPSDPPFYQCAEV 2754
|||||
DB 2821 DAPDYGVYKPEMDLATMEERVORRYEYKLTTEFVADMTKIFDNCRYNPSDPPFYQCAEV 2880
|||||
QY 2755 LESFFVQKLKFKGKASRSHNNKLQSTAS 2781
|||||
DB 2881 LESFFVQKLKFKGKASRSHNNKLQSTAS 2907
|||||
RESULT 3
ID AAO16418
XX AAO16418 standard; Protein; 2759 AA.
AC AAO16418;
XX
DT 10-APR-2003 (first entry)
XX
DE Human nucleic acid-associated protein (NAAP) - SEQ ID No 15.
DE
DE
KW Human; nucleic acid-associated protein; NAAP; arteriosclerosis;
KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;
KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
KW Crohn's disease; transgenic animal; animal model.
XX
OS Homo sapiens.
XX
PN WO2003000864-A2.
XX
XX 03-JAN-2003.
XX
PF 20-JUN-2002; 2002WO-US21179.
XX
PR 22-JUN-2001; 2001US-300518P.
PR 29-JUN-2001; 2001US-301787P.
PR 29-JUN-2001; 2001US-301792P.
PR 29-JUN-2001; 2001US-301892P.
PR 29-JUN-2001; 2001US-301893P.
PR 06-JUL-2001; 2001US-303405P.
PR 06-JUL-2001; 2001US-303442P.
PR 15-MAR-2002; 2002US-364438P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;
PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IJ;
PI Ramkumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn MR;
PI Borowsky ML, Yao MG, Wallia NK, Bandman O, Lal PG, Becha SD;
PI Lee SY, Richardson TW, Elliott VS, Luo W, Tang YT, Zebarjadian Y;
PI Lu Y;
XX
XX WPI; 2003-201420/19.
DR N-PSDB; AAL51568.
XX
XX New nucleic acid-associated proteins and polynucleotides, useful for
PT diagnosing, treating or preventing cell proliferative (e.g. cancer),
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
PT disorders (e.g. AIDS)
XX
XX Claim 1; Page 234-241; 312pp; English.
PS
XX The invention comprises the amino acid and coding sequences of human
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC the invention are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of NAAP, such as: cell proliferative
CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
CC or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia

CC or mental retardation); neurological disorders (e.g. Alzheimer's disease,
CC Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
CC (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
CC the invention are useful for creating transgenic animals to model human
CC disease. The present amino acid sequence represents a human nucleic acid-
CC associated protein of the invention.
XX
SQ Sequence 2759 AA:

Query Match 93.7%; Score 13428; DB 24; Length 2759;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 2625; Conservative 6; Mismatches 7; Indels 264; Gaps 3;

QY 1 MVSEEEEDGDAEETQDSEDEDEDEEDDDSDYPEEMEDDDDDASCTESSFRSHST 60
|||||
DB 1 MVSEEEEDGDAEETQDSEDEDEDEEDDDSDYPEEMEDDDDDASCTESSFRSHST 60
|||||
QY 61 YSSTGRRKPRVHRPRSPILPEFPKSSSEDLMPVNEHIMNVIAIYVLRNFTV 120
|||||
DB 61 YSSTGRRKPRVHRPRSPILPEFPKSSSEDLMPVNEHIMNVIAIYVLRNFTV 120
|||||
QY 121 LRLSPFRFDFCAALVSQEOCTLMAMHVVLKAVLREEDTSNTTFGPADLKDSVNSTLY 180
|||||
DB 121 LRLSPFRFDFCAALVSQEOCTLMAMHVVLKAVLREEDTSNTTFGPADLKDSVNSTLY 180
|||||
QY 181 FIDGMTWPEVLRVYCSDKEYHHVLPYQAEADYPYGPVENKIKVLQFLVDQFLTNIARE 240
|||||
DB 181 FIDGMTWPEVLRVYCSDKEYHHVLPYQAEADYPYGPVENKIKVLQFLVDQFLTNIARE 240
|||||
QY 241 ELMSEGVIOYDHCRCVCHKLGLLCCETCSAVYHLECVKPPLEVPEDQCEVCVAHV 300
|||||
DB 241 ELMSEGVIOYDHCRCVCHKLGLLCCETCSAVYHLECVKPPLEVPEDQCEVCVAHV 300
|||||
QY 301 PGVTDCVAEIQNKPYIRHEPIGYDRSRKRYWFLNRRLIIEEDTENENKKIWIYTKVQ 360
|||||
DB 301 PGVTDCVAEIQNKPYIRHEPIGYDRSRKRYWFLNRRLIIEEDTENENKKIWIYTKVQ 360
|||||
QY 361 LAELIDCLDKYWEAELCKILBEMREEIHRHMDITEDLTNKARGSNKSLAANEEILES 420
|||||
DB 361 LAELIDCLDKYWEAELCKILBEMREEIHRHMDITEDLTNKARGSNKSLAANEEILES 420
|||||
QY 421 IRAKGDIDNVKSPETEKNETENDSKDAENREFEQDSLEKSDDKTTPDDPDEQK 480
|||||
DB 421 IRAKGDIDNVKSPETEKNETENDSKDAENREFEQDSLEKSDDKTTPDDPDEQK 480
|||||
QY 481 SE----- 482
||
DB 481 SEPTTEVGDGKNSVSNLGDNTTNATSEETSPSEGRSPVGCLETDPDSSNMAEKKVASEL 540
|||||
QY 483 -----VGDFKSEKSNGLSESPGAGKSGSTRITRLRNPDPSKLSQLKSQVAAAHEA 537
|||||
DB 541 PQDVPVGDGFKSEKSNGLSESPGAGKSGSTRITRLRNPDPSKLSQLKSQVAAAHEA 600
|||||
QY 538 NKLFPKEGKEVLVYVNSQGEISRLSTKEVIMKGNINNYFKLGQEGKYRVYHNQYSTNSFAL 597
|||||
DB 601 NKLFPKEGKEVLVYVNSQGEISRLSTKEVIMKGNINNYFKLGQEGKYRVYHNQYSTNSFAL 660
|||||
QY 598 NKQHREDDHDKRRHLAHKFLTPAGBFKWNKSGVHSGKVLTIISLRLTITOLENNIPSSFL 657
|||||
DB 661 NKQHREDDHDKRRHLAHKFLTPAGBFKWNKSGVHSGKVLTIISLRLTITOLENNIPSSFL 720
|||||
QY 658 HPNWSHRANWIKAVQMCSPREFALALAILCAVKPVPVYMLPTWREFLGLRLHRTSTIE 717
|||||
DB 721 HPNWSHRANWIKAVQMCSPREFALALAILCAVKPVPVYMLPTWREFLGLRLHRTSTIE 780
|||||
QY 718 REEKEKVKKKEKQEBEETMQOATWVKYTFPVKHQVWKOKGGEYRTVGGGWSWISKTHV 777
|||||
DB 781 REEKEKVKKKEKQEBEETMQOATWVKYTFPVKHQVWKOKGGEYRTVGGGWSWISKTHV 840
|||||
QY 778 YRFVPKLPGNTVNVYRKSLGKTKNNNDENDESCKRSRPPKIKIEPDSKDEKVGSD 837
|||||
DB 841 YRFVPLPGNTVNVYRKSLGKTKNNNDENDESCKRSRPPKIKIEPDSKDEKVGSD 900
|||||

Qy	838	AAKADQNMEDISKITEKDQDYKELDDSDSRCKPEEPMEVDDDMKTESHVNCQESSQV	897
Db	901	AAKADQNMEDISKITEKDQDYKELDDSDSRCKPEEPMEVDDDMKTESHVNCQESSQV	960
Qy	898	DVNVSGFHLRYSYKKTKSKLDGLLERRIKQFTLEEKORLEKIKLEGIGIKGTST	957
Db	961	DVNVSGFHLRYSYKKTKSKLDGLLERRIKQFTLEEKORLEKIKLEGIGIKGTST	1020
Qy	958	NSSKNLSESPVITKAKGCGSDSMRQEQSPNANDOPEDLIQCSOSDSSVLKMSDPSHT	1017
Db	1021	NSSKNLSESPVITKAKGCGSDSMRQEQSPNANDOPEDLIQCSOSDSSVLKMSDPSHT	1080
Qy	1018	TNKLYPKDRVLDDVIRSPTKPKQNSIENDIEEKVSDIASRQEPKSKTKGNDFID	1077
Db	1081	TNKLYPKDRVLDDVIRSPTKPKQNSIENDIEEKVSDIASRQEPKSKTKGNDFID	1140
Qy	1078	DSKLASADDGTGLCKNKKPLIOESDTIVSSSKSALHSSVPKSTNDRDATPLSRAMDFE	1137
Db	1141	DSKLASADDGTGLCKNKKPLIOESDTIVSSSKSALHSSVPKSTNDRDATPLSRAMDFE	1200
Qy	1138	GKLGCDSESNSTLSSSDTVSIODSSSEDMIVQNSNESISEQFRTREDDVEVLEPLKCEL	1197
Db	1201	GKLGCDSESNSTLSSSDTVSIODSSSEDMIVQNSNESISEQFRTREDDVEVLEPLKCEL	1260
Qy	1198	VSGESTGNCEDRLPVKGTANGKPKSQOKKLEERPYNKCSQDKLKNNTDKKNNENRESE	1257
Db	1261	VSGESTGNCEDRLPVKGTANGKPKSQOKKLEERPYNKCSQDKLKNNTDKKNNENRESE	1320
Qy	1258	KKGQRTSTFOINGKDNPKIYLKGECLKEISESRVSGNVEPKVNNINKLIIPENDIKSLT	1317
Db	1321	KKGQRTSTFOINGKDNPKIYLKGECLKEISESRVSGNVEPKVNNINKLIIPENDIKSLT	1380
Qy	1318	VKESAIRPFINGDVIEMEDFNERNSESKSHLLSSSDAEGNYRDSLETLPSTKESDSTQTT	1377
Db	1381	VKESAIRPFINGDVIEMEDFNERNSESKSHLLSSSDAEGNYRDSLETLPSTKESDSTQTT	1440
Qy	1378	TPSASCPESNSVNOVEDMEIETSEVKVTSPTSEESNLNDFIDENGLPIPNKNENVN	1437
Db	1441	TPSASCPESNSVNOVEDMEIETSEVKVTSPTSEESNLNDFIDENGLPIPNKNENVN	1500
Qy	1438	GESKRKVTITVTWMTSTVATESKTVIKVEGDKQTVVSSSTENCAKSTVTTTTVTVKLS	1497
Db	1501	GESKRKVTITVTWMTSTVATESKTVIKVEGDKQTVVSSSTENCAKSTVTTTTVTVKLS	1560
Qy	1498	TPSTGGSDIISVKEQSKTVVTTVTDLSLTTGTGLTVTSMTVSKKEYSTRDKVKLMKFSRP	1557
Db	1561	TPSTGGSDIISVKEQSKTVVTTVTDLSLTTGTGLTVTSMTVSKKEYSTRDKVKLMKFSRP	1620
Qy	1558	KKTRSGTALPSYRKFTVKSTKSIIFVLPNDLKLARKGGIREVPYFNYNNAKPALEDWPY	1617
Db	1621	KKTRSGTALPSYRKFTVKSTKSIIFVLPNDLKLARKGGIREVPYFNYNNAKPALEDWPY	1680
Qy	1618	PSPRFTFGITWRYRLQTVKSLAGVSLMLRLWASLRWDDMAKVPVPGGGSTRTTSETI	1677
Db	1681	PSPRFTFGITWRYRLQTVKSLAGVSLMLRLWASLRWDDMAKVPVPGGGSTRTTSETI	1740
Qy	1678	TTTEIIRKRDVGPVIREYECIRKIICIPGVPEPTKPTPQRKGLRSSALRPKRPTPK	1737
Db	1741	TTTEIIRKRDVGPVIREYECIRKIICIPGVPEPTKPTPQRKGLRSSALRPKRPTPK	1800
Qy	1738	QTGPVITETWAAEELELEWIRAFARVEKEKAQAVEQOAKRLEQOQKPTVIATSTTSPT	1797
Db	1801	QTGPVITETWAAEELELEWIRAFARVEKEKAQAVEQOAKRLEQOQKPTVIATSTTSPT	1860
Qy	1798	SSTTSTISPAQKVMVAPISGVTGTMVLTKVGSPTATVTFQONKNFPHOTFATVWKQOQ	1857
Db	1861	SSTTSTISPAQKVMVAPISGVTGTMVLTKVGSPTATVTFQONKNFPHOTFATVWKQOQ	1920
Qy	1858	SNSGVVQVQKVLGIIPSTGTSQQTFTSFQPRATVTVIRNTSGSGGTTNSQVITGPQ	1917
Db	1921	SNSGVVQVQKVLGIIPSTGTSQQTFTSFQPRATVTVIRNTSGSGGTTNSQVITGPQ	1980
Qy	1918	IRPGMTVIRTPLQQSTGLKAIIRTPVMVPGAPQVMTQIIRGQPVSTAVSAPNTVSSSTP	1977

Db	1981	IRPGMTVIRTPLQQSTGLKAIIRTPVMVPGAPQVMTQIIRGQPVSTAVSAPNTVSSSTP	2040
Qy	1978	GOKSLTSATSTSNIQSSASOPPRPOGOVKLTMAOLTQLTQHGNGOGLTVVIQOGQTT	2037
Db	2041	GOKSLTSATSTSNIQSSASOPPRPOGOVKLTMAOLTQLTQHGNGOGLTVVIQOGQTT	2100
Qy	2038	GOLQLIPQGVTVLPFGQOOLMAAMPNGTVQRFLPTLATTTASTTTTSTVSTAAGTG	2097
Db	2101	GOLQLIPQGVTVLPFGQOOLMAAMPNGTVQRFLPTLATTTASTTTTSTVSTAAGTG	2160
Qy	2098	EQRSKLSFOMQVHQDKTLPPAQSSSYGAKAPOQAOPSAPOQTOPQSPAPQPEVQIQ	2157
Db	2161	EQRSKLSFOMQVHQDKTLPPAQSSSYGAKAPOQAOPSAPOQTOPQSPAPQPEVQIQ	2220
Qy	2158	PEVOTOTVSSHVPSEAQPTHAQSSKPKQVAAQSQPSNVQSGSPVRVQSPSTRIRPSTP	2217
Db	2221	PEVOTOTVSSHVPSEAQPTHAQSSKPKQVAAQSQPSNVQSGSPVRVQSPSTRIRPSTP	2280
Qy	2218	SQSPGQSQSOVQTTTQPIQPIHTSLQIPSGQSQSOQVQSQSTOTLSGGQTLNOVSYS	2277
Db	2281	SQSPGQSQSOVQTTTQPIQPIHTSLQIPSGQSQSOQVQSQSTOTLSGGQTLNOVSYS	2319
Qy	2278	SPSRPQLQIQOPQOQVYIAVPLQLOQOQVLSQIQSVVAQIQAAQSGVQPOQIKLQLPQIQ	2337
Db	2320	-----	2319
Qy	2338	QSSAVQTHQTONVTVQAASVQBLQVQOLRDQOQKQKQOQIEIKREHTLQASNOSEII	2397
Db	2320	-----	2319
Qy	2398	QKQVVMKHNNAVIEHLKQKKSMTFAERENORMIVCNQVMKYILDKIDKEKQAAKKRKE	2457
Db	2320	--QVVMKHNNAVIEHLKQKKSMTFAERENORMIVCNQVMKYILDKIDKEKQAAKKRKE	2377
Qy	2458	ESVEQKRSKONATKLSALLEFKHKEOLRAEILKRALLDKDLQIEVQOEELKRDLIKKEKD	2517
Db	2378	ESVEQKRSKONATKLSALLEFKHKEOLRAEILKRALLDKDLQIEVQOEELKRDLIKKEKD	2437
Qy	2518	LMQLAQATAVAAPCPVTVLPAPPAPPPPPPPGVQHTGLLSTPTPLPVASOKRKEEE	2577
Db	2438	LMQLAQATAVAAPCPVTVLPAPPAPPPPPPPGVQHTGLLSTPTPLPVASOKRKEEE	2497
Qy	2578	KDSSSKSKKKMISTTSKETKDTKLYCICKTPYDESK-----FYIGCDRCQNNYHGRCVGILQSEA	2615
Db	2498	KDSSSKSKKKMISTTSKETKDTKLYCICKTPYDESKFYIGCDRCQNNYHGRCVGITEK	2557
Qy	2616	-----FYIGCDRCQNNYHGRCVGILQSEA	2639
Db	2558	EAKKMDVYICNDCKRAQSGSSEELYCICRTPYDESQFYIGCDRCQNNYHGRCVGILQSEA	2617
Qy	2640	ELIDEEYVCPQOQSTEDAMTVLPTLEKDYGLKRVLSLQAHKMAWPFLEVPDNDAPDY	2699
Db	2618	ELIDEEYVCPQOQSTEDAMTVLPTLEKDYGLKRVLSLQAHKMAWPFLEVPDNDAPDY	2677
Qy	2700	YGVILKEPMDLATMEERVORRYEKLTEFVADMTKIFDNCRYNPSDSPFFYQCAEVLBSFF	2759
Db	2678	YGVILKEPMDLATMEERVORRYEKLTEFVADMTKIFDNCRYNPSDSPFFYQCAEVLBSFF	2737
Qy	2760	VQKLKGFKASRSHNNKLOSTAS	2781
Db	2738	VQKLKGFKASRSHNNKLOSTAS	2759

RESULT 4
AAB94078
ID AAB94078 standard; Protein; 557 AA.
XX
AC AAB94078;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:14273.

CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.

XX
SQ Sequence 328 AA;
Query Match 10.3%; Score 1474; DB 21; Length 328;
Best Local Similarity 94.8%; Pred. No. 6.1e-68;
Matches 289; Conservative 0; Mismatches 14; Indels 2; Gaps 1;

QY 406 NKFLAANFEILIESIRAKGDDIDNVKSP--EETEDKNETENDSKDAEKREFEQSL 463
Db 12 NVSLSGGTNEILIESIRAKGDDIDNVKSPGTEETEDKNETENDSKDAEKREFEQSL 71
QY 464 EKSDDKTDPDDPEQKSEVGVDFKSEKNGELSESPGAGSGSTRITRLRNPDSKLS 523
Db 72 EKSDDKTDPDDPEQKSEVGVDFKSEKNGELSESPGAGSGSTRITRLRNPDSKLS 131
QY 524 QLKSOQVAAAHAANKLFEGKEVLVNVSGEISRLSTKKEVTMKGINNINNYFKLGQEGY 593
Db 132 QLKSOQVAAAHAANKLFEGKEVLVNVSGEISRLSTKXKXVXMKGINNINNYFKLGQEGY 191
QY 584 RVYHNOYSTNSPALNKHQREDHDKRHLAHKFLTPAGEFKWNGSVHGSKVLTITLRL 643
Db 192 RVYHNOYSTNSPALNKHQREDHDKRHLAHKFLTPAGEFKWNGSVHGSKVLTITLRL 251
QY 644 TITQLENNIPSSPLHPNASHRANWIKAVQCMCKPREFALALAILCAVKPVMVLMPIWRE 703
Db 252 TITQLENNIPSSPLHPNASHRANWIKAVQCMCKPREFALALAILCAVKPVMVLMPIWRE 311
QY 704 FLIGHT 708
Db 312 SLIGHT 316

RESULT 7
ABU70846
ID ABU70846 standard; Protein; 191 AA.
XX AC ABU70846;
XX DT 10-JUN-2003 (first entry)
XX DE Human adipocyte Selected Interacting domain, SID, #477.
XX KW Human; prev; adipocyte; SID; selected interacting domain;
KW anorectic; antidiabetic; protein-protein interaction; diabetes;
KW yeast 2-hybrid assay; metabolic disorder; obesity.
XX OS Homo sapiens.
XX PN WO200286122-A2.
XX PD 31-OCT-2002.
XX PF 14-MAR-2002; 2002WO-EP03768.
XX PR 14-MAR-2001; 2001US-275734P.
XX PA (HYBR-) HYBRIGENICS.
XX PI Legrain P, Daviet L;
XX DR WPI: 2003-103412/09.
XX DR N-PSDB; ACA57390.
XX PT New complex between two interacting proteins in adipocyte cells, useful
PT for identifying selected interacting domains that modulate protein
PT interactions, or for preventing or treating metabolic disorders such as
PT obesity or diabetes
XX PS Claim 6; Page 267-268; 382pp; English.

XX The invention relates to a complex between two interacting proteins in
CC adipocyte cells, given in the specification. The proteins are identified
CC by selecting a bait protein from a known adipocyte marker and then
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC members of an adipocyte cDNA library. The proteins are designated SID
CC (RTM) (selected interacting domains) proteins. Also included are a
CC polynucleotide encoding a polypeptide in the adipocyte cells, a
CC recombinant host cell expressing at least one of the interacting
CC polypeptides of the complex, selecting a modulating compound in adipocyte
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
CC sequences given in the specification (including its fragment or variant),
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
CC given in the specification (including its fragment or variant), a vector
CC comprising the SID (RTM) polynucleotide, a recombinant host cell
CC comprising the vector, a protein chip comprising the polypeptides and
CC a record comprising all or part of the data, listed in the specification.
CC The complex, polypeptides, polynucleotides and compounds are
CC useful for preventing or treating metabolic disorders such as obesity
CC or diabetes. The polynucleotides are useful as probes or primers. The
CC complex is particularly useful for identifying selected interacting
CC domains (SID (RTM)) for screening drugs that modulate the protein
CC interaction, thus exhibiting the therapeutic effect. The present
CC sequence represents a SID (prey) protein of the invention.

XX
SQ Sequence 191 AA;
Query Match 7.2%; Score 1030; DB 24; Length 191;
Best Local Similarity 100.0%; Pred. No. 2e-45; Indels 0; Gaps 0;
Matches 191; Conservative 0; Mismatches 0;

QY 2556 HTGLLSTPTLPVASKRKRREEDSSSKKKKKMISTTSKTKKDKLYCICKTPYDESK 2615
Db 1 HTGLLSTPTLPVASKRKRREEDSSSKKKKKMISTTSKTKKDKLYCICKTPYDESK 60
QY 2616 FYIGCDRCQNWYHRCVGLQSEAEILDEYVCPQCOSTEDAMTVLPLEKDEGLKRVL 2675
Db 61 FYIGCDRCQNWYHRCVGLQSEAEILDEYVCPQCOSTEDAMTVLPLEKDEGLKRVL 120
QY 2676 RSLQAHKMAWPFLEPVDNDADPDYGVYKPEMDLATMEERVQRRYKELTEFVADMTKIF 2735
Db 121 RSLQAHKMAWPFLEPVDNDADPDYGVYKPEMDLATMEERVQRRYKELTEFVADMTKIF 180
QY 2736 DNCYNPSDS 2746
Db 181 DNCYNPSDS 191

RESULT 8
AAU04323
ID AAU04323 standard; Protein; 238 AA.
XX AC AAU04323;
XX DT 18-JUN-1999 (first entry)
XX DE Human secreted protein SEQ ID NO:69.
XX KW Human; secreted protein; cancer; tumour; developmental abnormality;
KW foetal deficiency; blood disorder; immune system disorder; inflammation;
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW digestive disorder; endocrine disorder; infection; AIDS.
XX OS Homo sapiens.
XX PN WO9910363-A1.
XX PD 04-MAR-1999.
XX PF 27-AUG-1998; 98WO-US17709.

```
PR 29-AUG-1997; 97US-0056271.
PR 29-AUG-1997; 97US-0056073.
PR 29-AUG-1997; 97US-0056247.
PR 29-AUG-1997; 97US-0056270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Fan P, Kyaw H, Rosen CA, Ruben SM, Wei YF;
XX WPI; 1999-190585/16.
XX
XX New isolated human genes and the secreted polypeptides they encode
XX - useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders
XX
XX Disclosure; Page 18; 170pp; English.
XX
XX AAX30145 to AAX30173 represent 29 isolated human secreted protein genes.
XX AAY04293 to AAY04321 represent the secreted proteins encoded by the 29
XX human genes. The genes and their corresponding secreted polypeptides are
XX useful for preventing, treating or ameliorating medical conditions,
XX e.g. by protein or gene therapy. Also pathological conditions can be
XX diagnosed by determining the amount of the new polypeptides in a sample
XX or by determining the presence of mutations in the new genes. Specific
XX uses are described for each of the 29 genes, based on which tissues they
XX are most highly expressed in, and include developing products for the
XX diagnosis or treatment of cancer, tumours, developmental abnormalities
XX and foetal deficiencies, blood disorders, diseases of the immune system,
XX autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
XX disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
XX disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
XX disorders, digestive/endocrine disorders, infections and AIDS. The
XX polypeptides are also useful for identifying their binding partners.
XX The sequences given in AAX30174 to AAX30182 and AAY04322 to AAY04334 are
XX used in the exemplification of the present invention.
XX
XX Sequence 238 AA;
XX
XX Query Match 6.9%; Score 982.5; DB 20; Length 238;
XX Best Local Similarity 78.8%; Pred. No. 7.8e-43;
XX Matches 189; Conservative 7; Mismatches 19; Indels 25; Gaps 3;
XX
QY 2563 PTLVASOKRREK-----DSSSKSKKKMKMISTTSKTKK-----DT 2601
DQ ||||| :||| : : : ||||
DQ 3 PTPR----KTPYDESKFYIGCDLCTNWHGCVGITEKAKMDVYICNCKKRAQGSSE 58
QY 2602 KLYCICKTPYDESKFYIGCDRCQNWYHGRVCGVILQSEAEILDEVVCPQCQSTEDAMTVLT 2661
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DQ 59 ELYCICRTPYDESOFYIGCDRCQNWYHGRVCGVILQSEAEILDEVVCPQCQSTEDAMTVLT 118
QY 2662 PLTEKDYGLKRVLSLQAKHMAWPFLEPVDNDAPDYGVYIKPEMDLATMEERVQRRYY 2721
DQ |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DQ 119 PLTEKDYGLKRVLSLQAKHMAWPFLEPVDNDAPDYGVYIKPEMDLATMEERVQRRYY 178
QY 2722 EKLTEFVADMKIFDNCRYNPSDSPFYQCAEVLESFVQKLGFKASRSNNKLOSTAS 2781
DQ |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DQ 179 EKLTEFVADMKIFDNCRYNPSDSPFYQCAEVLESFVQKLGFKASRSNNKLOSTAS 238
XX
RESULT 9
AAU19446
ID AAU19446 standard; Protein; 237 AA.
XX
XX AAU19446;
XX
XX 04-DEC-2001 (first entry)
XX
XX Human diagnostic and therapeutic polypeptide (DITHP) #32.
XX
XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
XX cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
XX acquired immune deficiency syndrome; AIDS; autoimmune disorder;
XX respiratory disorder.
```

```
XX OS Homo sapiens.
XX XX WO200162927-A2.
XX PN
XX PD 30-AUG-2001.
XX
XX 21-FEB-2001; 2001WO-US06059.
XX PF
XX 24-FEB-2000; 2000US-0184693.
XX PR 24-FEB-2000; 2000US-0184697.
XX PR 24-FEB-2000; 2000US-0184698.
XX PR 24-FEB-2000; 2000US-0184768.
XX PR 24-FEB-2000; 2000US-0184769.
XX PR 24-FEB-2000; 2000US-0184770.
XX PR 24-FEB-2000; 2000US-0184771.
XX PR 24-FEB-2000; 2000US-0184772.
XX PR 24-FEB-2000; 2000US-0184773.
XX PR 24-FEB-2000; 2000US-0184774.
XX PR 24-FEB-2000; 2000US-0184776.
XX PR 24-FEB-2000; 2000US-0184777.
XX PR 24-FEB-2000; 2000US-0184797.
XX PR 24-FEB-2000; 2000US-0184813.
XX PR 24-FEB-2000; 2000US-0184837.
XX PR 24-FEB-2000; 2000US-0184841.
XX PR 24-FEB-2000; 2000US-0185213.
XX PR 24-FEB-2000; 2000US-0185216.
XX PR 12-MAY-2000; 2000US-0203785.
XX PR 15-MAY-2000; 2000US-0204226.
XX PR 16-MAY-2000; 2000US-0204525.
XX PR 16-MAY-2000; 2000US-0204821.
XX PR 16-MAY-2000; 2000US-0204908.
XX PR 16-MAY-2000; 2000US-0205232.
XX PR 17-MAY-2000; 2000US-0204815.
XX PR 17-MAY-2000; 2000US-0204863.
XX PR 17-MAY-2000; 2000US-0205221.
XX PR 17-MAY-2000; 2000US-0205285.
XX PR 17-MAY-2000; 2000US-0205286.
XX PR 17-MAY-2000; 2000US-0205287.
XX PR 17-MAY-2000; 2000US-0205323.
XX PR 17-MAY-2000; 2000US-0205324.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
XX Chen A, D'sa SA, Amshey S, Dahl CR, Dam TC, Daniels SE;
XX DuFour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL;
XX Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
XX Wright RJ, Yap PE, Yu JW, Bradley DL, Bratcher SR, Chen W;
XX Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX WPI; 2001-502867/55.
XX N-PSDB; AAS31017.
XX
XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
XX enzymes, hormones and receptors, useful in diagnostics and therapeutics
XX
XX Claim 27; Page 418-419; 522pp; English.
XX
XX The invention relates to polynucleotides (I) encoding diagnostic and
XX therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
XX and proteins involved in growth and development and receptors. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate DITHP expression. For example, (I) and
XX (II) may be used to treat disorders associated with decreased polypeptide
XX expression by rectifying mutations or deletions in a patient's genome,
XX that affect the activity of the DITHPs, by expressing inactive proteins
XX or supplementing the patient's own production of them. (I) and (II)
XX may be used to treat diseases, for example, cell proliferative disorder,
XX Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
XX leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
XX (I) may be used to produce the DITHPs, by inserting the nucleic acids
```

CC into a host cell and culturing the cell to express the protein. (I) and
 CC its complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. (II) may also be used as antigens in the production of
 CC antibodies against DITHPs and in assays to identify modulators of DITHP
 CC expression and activity. The anti-DITHP antibodies and antagonists may
 CC also be used to down regulate expression and activity. The anti-DITHP
 CC antibodies may also be used as diagnostic agents for detecting the
 CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbant
 CC assay (ELISA)). AAU19415-AAU19625 represent human diagnostic and
 CC therapeutic (DITHP) polypeptides of the invention.

XX
 SQ Sequence 237 AA;

Query Match 6.8%; Score 981; DB 22; Length 237;
 Best Local Similarity 87.5%; Pred. No. 9.2e-43;
 Matches 182; Conservative 4; Mismatches 6; Indels 16; Gaps 1;

QY 2590 ISTSTETKK-----DTKLYCICKTPYDESKFYIGCDRCQNNYHGRGVG 2633
 DB 30 VGITEAKKMDVYICNDCKRAQGSSEELYCICRTPYDESQFYIGCDRCQNNYHGRGVG 89

QY 2634 ILQSEALIDEYVCPQCQSTEDAMTVLPTLTKDYEGLRKRLSLQAHKMAWPFLEPVD 2693
 DB 90 ILQSEALIDEYVCPQCQSTEDAMTVLPTLTKDYEGLRKRLSLQAHKMAWPFLEPVD 149

QY 2694 NDAPDYGVYIKPEMDLATMEERVQRRYIEKLTFFVADMTKIFDNCRYNPSDSPFFYQCAE 2753
 DB 150 NDAPDYGVYIKPEMDLATMEERVQRRYIEKLTFFVADMTKIFDNCRYNPSDSPFFYQCAE 209

QY 2754 VLESFFVQKLGKFKASRSHNNKLOSTAS 2781
 DB 210 VLESFFVQKLGKFKASRSHNNKLOSTAS 237

RESULT 10
 AAU78845
 ID AAU78845 standard; Protein; 149 AA.
 AC AAU78845;
 DT 18-JUN-2002 (first entry)
 XX
 DE SCAN/KRAB protein SKAT-2 binding protein, 7A1.
 XX
 KW SCAN/KRAB protein associated with a Th2 phenotype; SKAT-2; asthma;
 KW allergy; hay fever; atopic dermatitis; allergic rhinitis;
 KW Alzheimer's disease; neurological disorder; apoptosis;
 KW T-helper response; vaccination; immune response; autoimmune disease;
 KW infection; cancer; kruppel-like zinc finger protein; 7A1; human.
 XX
 OS Homo sapiens.
 XX
 PN GB2364051-A.
 XX
 PD 16-JAN-2002.
 XX
 PF 06-APR-2000; 2000GB-0008549.
 XX
 PR 06-APR-2000; 2000GB-0008549.
 XX
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI Blanchard AD, Champion B, Page KR;
 XX
 DR WPI; 2002-306794/35.
 XX
 DR N-PSDB; ABK47549.
 XX
 XX New SKAT-2 polypeptide transcription factor, useful for treating or
 PT preventing e.g. allergy and Alzheimer's disease, also related nucleic
 PT acid, antibodies and modulators -
 XX

PS
 XX Disclosure; Page 37-38; 47pp; English.

CC The invention describes an isolated SCAN/KRAB protein associated with a
 CC Th2 phenotype (SKAT-2) polypeptide (I). (I), especially where expressed
 CC in recombinant cells, is used to identify agents that modulate SKAT-2
 CC activity, and may be useful therapeutically. The agents, also (I) and
 CC the polynucleotides (II) that encode it, are useful for treatment and
 CC prevention of diseases responsive to SKAT-2 modulation, specifically
 CC asthma; allergies (hay fever, atopic dermatitis; allergic rhinitis etc.);
 CC Alzheimer's disease (and other neurological disorders, particularly where
 CC related to apoptosis); conditions involving T-helper responses, and
 CC apoptosis; also in vaccination against an antigen to enhance/repress Th2
 CC or humoral responses, to reduce development of a Th1 phenotype, and to
 CC manipulate the immune response in autoimmune diseases, infections and
 CC cancer. (II) is also used for recombinant production of (I); as source
 CC of antisense therapeutics and as primers and probes, e.g. for diagnostic
 CC detection of mutations and for monitoring SKAT-2 expression in
 CC association with disease. This is the amino acid sequence of the peptide
 CC 7A1, that binds specifically to the kruppel-like zinc finger protein,
 CC SKAT-2.

XX
 SQ Sequence 149 AA;

Query Match 5.5%; Score 789; DB 23; Length 149;
 Best Local Similarity 100.0%; Pred. No. 3.7e-33;
 Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2633 GILQSEALIDEYVCPQCQSTEDAMTVLPTLTKDYEGLRKRLSLQAHKMAWPFLEPVD 2692
 DB 1 GILQSEALIDEYVCPQCQSTEDAMTVLPTLTKDYEGLRKRLSLQAHKMAWPFLEPVD 60

QY 2693 PDAPDYGVYIKPEMDLATMEERVQRRYIEKLTFFVADMTKIFDNCRYNPSDSPFFYQCA 2752
 DB 61 PDAPDYGVYIKPEMDLATMEERVQRRYIEKLTFFVADMTKIFDNCRYNPSDSPFFYQCA 120

QY 2753 EVLESFFVQKLGKFKASRSHNNKLOSTAS 2781
 DB 121 EVLESFFVQKLGKFKASRSHNNKLOSTAS 149

RESULT 11
 AAU31944
 ID AAU31944 standard; Protein; 246 AA.
 AC AAU31944;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #2435.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-611725/70.
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -

XX PS Claim 20; Page 529-530; 765pp; English.

XX CC The invention relates to novel human secreted polypeptides. The

CC polypeptides and antibodies to the polypeptides are useful for

CC determining the presence of or predisposition to a disease associated

CC with altered levels of polypeptide. The polypeptides are also useful for

CC identifying agents (agonists and antagonists) that bind to them. Cells

CC expressing the proteins are useful for identifying a therapeutic agent

CC for use in treatment of a pathology related to aberrant expression or

CC physiological interactions of the polypeptide. Vectors comprising

CC the nucleic acids encoding the polypeptides and cells genetically

CC engineered to express them are also useful for producing the proteins.

CC The proteins are useful in genetic vaccination, testing and

CC therapy, and can be used as nutritional supplements. They may be used to

CC increase stem cell proliferation; to regulate haematopoiesis; and in

CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;

CC immune suppression and/or stimulation; as anti-inflammatory agents; and

CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid

CC sequences of novel human secreted proteins of the invention.

XX Sequence 246 AA;

Query Match 5.4%; Score 781; DB 22; Length 246;

Best Local Similarity 67.8%; Pred. No. 1.9e-32;

Matches 162; Conservative 10; Mismatches 63; Indels 4; Gaps 2;

QY 751 HOVWKQGEYRVYGYGWSWISKTHYRVFPKLPNGTNNYRKSLGEGTKNNMDENDES 810

DB 7 HOVWKQGEYRVYGYGWSWISKTHYRVFPKLPNGTNNYRKSLGEGTKNNMDENDES 66

QY 811 DKRCRSRPPKIKTEPDEKDEKGVDAAGADQNMDSIKITEKDDQYKELLDSDSK 870

DB 67 DKRCRSRPPKIKTEPDEKDEKGVDAAGADQNMDSIKITEKDDQYKELLDSDSK 126

QY 871 PKCEPMEVDDMTESHVNCQESSQVD--VVNVSEGFHLRTSYKKTKSKLDGLLRR 928

DB 127 PKCERTN--GSRRHENHRVTCKLSEFSSRGCCXGFSKDXLQKENKIIQTRWTSXKE 184

QY 929 IKQFTLEBKQRIKLEGGIKGKTSTNSKNLSESPVITKAKEGQSDSMRQEQSP 987

DB 185 NXYVTGRTATRNQVGGWXXGYRKDSTNSKNLSESPVITKAKEGQSDSDETRTEP 243

RESULT 12

ABB62887

ID ABB62887 standard; Protein; 997 AA.

XX AC ABB62887;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 15453.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX OS pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO2001171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL06990.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX PS Disclosure; SEQ ID NO 15453; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA

CC sequences (ABLI01840-ABLI16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 997 AA;

Query Match 4.7%; Score 672.5; DB 22; Length 997;

Best Local Similarity 24.7%; Pred. No. 5.3e-26;

Matches 311; Conservative 145; Mismatches 395; Indels 409; Gaps 51;

QY 1394 DMEIETSEVKVTSSPITSEESNL-----SNDFID---ENGLPINK 1432

DB 6 NVEIESTEDSIVTCLNSGNAEDVDMTPGWRKRKNOKSKSYIGTKDVLDTLDKDIPLNK 65

QY 1433 NENVNGESKRTVITEVTMTSIVAVESKTVIKVEKGDQKOTVVSSSTENCAKSVTTTTT 1492

DB 66 -----QNRFPIT-----ARPVREC-----VKYERETFENGNERV----- 97

QY 1493 VTKLSTPSTGGSV-----DIISVKEQSKTVVTTTDTSLTTTGGTFLVTMTVSKEYSTRDK 1548

DB 98 ----YSTSSPGRVYLLNDAAKLYEQA-----VKTEDK 126

QY 1549 VKLMKFSRPKKTSGTALPSYRFPVTKSTKTSFVLPNDLLKILARKGGIREVPYFNNA 1608

DB 127 STITK--KPSYSR---YPLISNPLTHKKRSLVLPFRFELLKILARLGKSSNNGFHHAA 180

QY 1609 KPALDIWYPSRPFEGITWRYRLQTVKSLAGVSLMLRLLWASLRDWDMAAKVPVGGST 1668

DB 181 KNN-TIWQYQCSRLPFTCWSYTSNATSLSSALQLRLWCLRWDDMIK-PPSTDGK 238

QY 1669 RTSETETEITTEIKRRDVGYPYIGRFEYCIRKICPIGVPEPKETPTPQRKGLRSAL 1728

DB 239 HQVTTDTTEIVTLELLKLRHSGRYGKTSYLRRKVVIPLEMPKTREV-TSIRSLR---- 293

QY 1729 RPKRPETPKQGPVLIETWVAEELELWEIPAFARVEKEKAQAVEQ--AKKRLQEQKPT 1787

DB 294 KKRKRAESQPTEPQITEEWDDEDEKLELWEIKFMGEKQEKAKLSAVTSVRSARQLE----- 348

QY 1788 VIATSTTSPSTSTISPAQKVMVAP--IGSVTTGTWKVLTITKVGSPATVTFQONKFNH 1846

DB 349 --ASGNSGNTSTNGALGVAGRVQLAPKLSDDVK--EKMEQQLXL---QRAVHQQRK--- 398

QY 1847 QTFATVWKQGSNSGVVQVQOKVLG-----IIPSTGTSGQFFTSFQPTATVTRPNTSG 1902

DB 399 ----LVATGEITRSVTPVKQGVIGSRVIVKNPDGTR----- 432

QY 1903 SGGTTSNSQVITGPQIRPGMTVIRTPLOOSTLGRKAITRTPVWVQPGAPQVMTQIIRGP 1962

DB 433 -----II-----QQAVTQVSRGTG 446

QY 1963 VSTAVSAPNTVSTPGOKSLTSATSTNSIOSASQPPRPOQGVKLTMALQTQLTQHGCG 2022

DB 447 ANTAATAA---ASPTVG-----GSTSQS-NFSTSTPHKVQ----- 477

QY 2023 NOGTLTVIQOGQGTGQLIPIQGVTVLPG--PGQQLMOAAMPNGTVQRFLETPLATAT 2080

DB 478 -----IIRG-----PDGKVSVRGLNPGQQLVQ--MPDGKLT----- 506

Db 778 ATEVPAEEVEDEFAKP-----TPIAAEEBPIAGTPIPTDGIS-----GEEBIVK 822
Qy 724 VKKKEKKQBEETWQAOTWYKTFPPVKHVKWQKQGEERYVTGCGHWSWISKTHVYFVKP 783
Db 823 GTTPQTLIEQOEISESTEVE-----PVAED-----D 848
Qy 784 LPGNTNVN-YRKSLEGTKNNMDNDESKRCKSRPKKIKIBPDSEKD-----831
Db 849 LSSSTASATASSTEGVQDAASETTSAPARAGDKDEAATVTAQDKDDEVEQDATDL 908
Qy 832 ---EVKGSDAAGADQNMEDISKITEKKQODVKELLDSGCKPCKEEPMVEVDDMK-----884
Db 909 PVEDVQSTTAKTTTTPQKESSTEAEDAIEVTTSSPADKQOEVPPEAEAPADKHDEED 968
Qy 885 -----TESHVNCQESSQVDVNVNVEGFHL-RTSYKKKTKSSKLDGL 924
Db 969 VQTATDLPKISDGIQPPVVEATGQPETDEATDKPPSVLPPVSVQVPSSTAKVDNR 1028
Qy 925 LERRIKOFTL-----BEKORLEKIKLEGIGK-----IGKTSNKKNLSESPVITKA 972
Db 1029 NDFETEXTLPPSGEDOSSEPLPAMDLPAGIPGEGDCLVEGKTYANNTIVPATAPCDVSC 1088
Qy 973 KEGQCSMRQEQSGPNANNDOPE-----DLIOGCSQS-----DSVLKMSDPSTHNNKLYP 1023
Db 1089 K--CISSILVACQOOMEKLPENLEKCTVAADLLDCCCTYICDESTESAERKDEESTAK--P 1144
Qy 1024 KDRVLDDVIRSPTKCPKQNSIENDTEEK-----VSDLASRGQEPKTSKTK 1070
Db 1145 DNKIDEDVSELTE-EIPKDVIMPTGTEQPLSHVKPDEELQPVTSVPAQFDESTAKVD 1203
Qy 1071 GNDFIDDSKLASADDITGLCKNKPLIQPESDIIVSSKSLHSHVPSKSTNDRDATPL 1130
Db 1204 KKP--IDES-----AEDKKPIGESEED-----SKPIDESEDEKKPV 1237
Qy 1131 SRAMDFEGLCDSESNSTLENSDTSVIQDSSSEDMIV-----QNSNESISEQFRT 1183
Db 1238 E-----ESAEDKKPVEDSEKEKPLPTVTPASETEKESKPEDEKKT 1278
Qy 1184 EQD-----VEVLEPLKCELVGSESTGNCEDRLPVKGTANGK---KPSQOKKLEERPYNKC 1236
Db 1279 EADFAAPTEQFATTPAQIADTAKEVDDKLATTSAPVSGEDELKPADEKKRTE-----1332
Qy 1237 SDQI---KLKNTDKNNENR-----ESEKKGORTST--FQINGKDNKPKIYLKCECLKEI 1287
Db 1333 TAQIPDAEIPASTDEPESSTELPTVDLDKKPEEDSTKTEAPESDKVPEVPTSASTENEI 1392
Qy 1288 SESRVVSGNVEPKVNNINKIIP--ENDIKSLTVKE-----SAIRPFINGDVI-----1332
Db 1393 EESDKFTTVAPPKISADETEPTAEEDLVPAITFPIESEFVETKPKFVQGPPLTLAPA 1452
Qy 1333 -MEDFNERNSETKSHLLSSSDAEGNVRDSLETLPSTKESDSTOTTP--SASCPSNSVN 1390
Db 1453 OPEKKPVDAAETSTADISTEPSEAEVEKASGETSESDNEIDAGASSIPVPVSADEDKTPS 1512
Qy 1391 QVEDMEITSEVKVVTSSPIT-SEESH---LSNDFIDENGLPI-----KNENNVGE 1439
Db 1513 TEKTEVE---ADDKFTVAPLAGDEESNLKPLQDIFEEEA-PVAVTTAAPSCKD---GE 1565
Qy 1440 SK-----RKTVITEVMTSTVATESKTVIKVEKGDQTVVSVSTENCAKSTVT 1487
Db 1566 QKPEVEEKPIEDGQKPIEDETSPTPS-----SENEIEPESDRATTIAPSKEE--PSEPS 1618
Qy 1488 TTTTITVTKLSTPSTGSGVDIISVREBQSKTVVTTTVDLSLTTTGGTLVTSMTVSKYSTRD 1547
Db 1619 TCAPTKEPAEPSTDAPESDS-KETPESEVPTIVAPA---GEKIPTSSITPDEEPTAT 1673
Qy 1548 KVKLMKFRPKTRSGTALPSYRKVFTKSTKSFVLPNDDLKKLARKGGIREVPYFNYN 1607
Db 1674 SAPVAKPDDEVEKETSTSEIPTDAPASSEEDNS-----STDQIPS-----EVP-----1717
Qy 1608 AKPALDIWYPSPRPTFGITWRYRLQTVKSLAGVSLMLRLLWASLRWDDMAAKYPPGG--1665
1718 KKP-----ETPAQTPEEGDI 1732
1666 -GSTRDETSEITTEITTEILIKRRDVGPIGRFEYCIKILCIPGVPETPKETTPORK-GL 1723
1733 VGATAAPTTSDEVPVPVORLPE-----EVLAIEIPOPSTETGI 1768
1724 RSALRPKRPEPKQGTGPIIETWVAEELELWEIRAFARVEK--EKAQAVEO-----O 1776
1769 KQODETTAAPSIDRK-----EPYVTEIDEAATTVAPISEKDEKPTTEEEKPEVKPTGEE 1822
1777 AKRLRQOQPTVIATSTSTSTTISPAQKVMVAPISGVTGTCKMLVTTKVGSPAT 1836
1823 PSEEEKKAPIEDOVSTEGPSTVEASEAGSTES-----SEVKPSTEGEVAEKPED---1873
1837 VTFQONKNFHQTFATVWKQGSNS-GVWVQVQKVLIIPSSSTG---TSQOTFTSFOPRTA 1892
1874 -----KOPSSTAQAPVETIPEISTELPAQDGDKPTSEAPVSDSDTSTA 1916
1893 TVTIR-PNTSGSGGTTNSQVITGPOIRPCMTVIRTPLOOSTLGLKALIRTPVMVOPGAQ 1951
1917 PSDEKIPSVSG-----EEVEGPEV-----TTASPOAAEDELKTPABSESSTD 1960
1952 QV-MTOIIRGOPYSTAVSAPNTVSTPGOKSLTSATST-----SNIQSSASQPPRQOQG 2005
1961 KVPETEYQKPEDETKADETPESVTQVSDVATSTAPVAGGDIKDEQAATTASP--EEEE 2018
2006 VKLTMALQTLQTOGHGNGOGLTVVIOGQOTTGOLQILPOGVTVLP--GPGQOLMQ-AAM 2062
2019 IKPTIAPAAEIPQ-----PSEKPEVDE-QEVESGTRKATPAESDQGPIDEIAPA 2065
2063 PNGTVQRFLLPTLATTATTATTTTSTTAAGTGEQRQSKLSPQMVQHDK-----TL 2116
2066 TSGPIDB-----ASTAAPTKEESTIVASAA-----SP-AVHDEILKDVTTQ 2106
2117 PPAQSSVSGPAKAQPOTA-----QPSAR-PQOTQOPSPAQPEVOTQPEVOT--OTTV 2166
2107 PVADEKEVAAPQDETKTISIDVSTDSPTAQDDEKQDKTEAPVAPTIVSSPTADSAADSSTP 2166
2167 SSHVPS-----EAQPTH---AQSSKQVA-AQSOPOSNOVGOSPVV--QSPSOTRIRPS- 2215
2167 TVEVPSVEIDTKPMDDIMSOTIAPHTADGAATSTEDQAPVTVSPQDAEKTPVSPAP 2226
2216 -----TPSOLSPGQOQSVQTTTQO-----PIPIQHTSLQIPISQGOPOQOVSQSTOT 2264
2227 QDSKPTSSRAPQDADEIPATATPLDDNKIPATVAPQTDGVPATAAPLDEDKIQTAAAP 2286
2265 L-----SSGOTLNQVSVSPSRPQLIOQOPQOVIAPVQLOQOQVLSQISQVVAIQ 2318
2287 LDEEKIPSTAAPLDDEKIPAPVSPVDFVEPSEKPAVSEYDGE-----ESTEPVVDHVE 2341
2319 AQSGVPOQIKLQLPQIOQSSAVQTHQIONVTVQAA-----SVQEQLOQVQOLRQOQ 2373
2342 TSTDEPTSDAKLKPTSPATPPSPESPAEIVPETAPELEKEVPEKATEQPELEKETP 2401
2374 KKKOQOIKREHPTLOASNOSEI-----IQKQVMKHNNAVIEHLKOKKSMTP 2420
2402 EKATEQPELEKETPEKATEQPELEKETPEKATEQPELEKETPEKATEQPEVDEKTTPEP 2461
2421 AER-----ERNQRMIVCNQVMKYLDKIDKEEQAAKKRREESVQKRSQKQATKLSA 2474
2462 VVKPSLSDSTEDE-----ESVESEESADKKDKNKETEEDTKKHPEEPVPA 2508
2475 L-----LFXKHKEQLRAELTKRALLDLQLOVEQFELKRLDKIKKEK 2516
2509 VVSEIPOPSEAVPTTGHLPFH---LASSTTTPPAVDDR-----VQEE-----DEEN 2553
2517 DLMOLAQATAVAACPVPVT-----PVLPAAPPAPPPPPPGVQHTGLLSTPTLTPVASOKR 2572
2554 TTVKLSSTTTSTTSPVTSAPSTTTVASOQQQPIITPPPYG--HA-----PEYDEY 2603
2573 KREBEKSSSKKSKKMMISTTSKTKKDTKLYCICKTPYDESKFYICGDCRQWNYHRCV 2632
2604 DEEEVFGTCRYAGKLYVSAQQIPRODDPCDFCFC-----FRSDII 2644

3759	Db	I- - - - - P L O K M T P I O V P H H P T - - - - - I I S K V V T V Q Q A T Q S V A S S P P L G S - - - - - L P P H K 3805
2109	Qy	Q V H O D K T L P P A Q S S V G P A K A Q P T - A O P S A R P Q P O T Q P S P A Q E V Q T O P E V Q T O T T V S 2167
3806	Db	N V H - - - - - L N A H Q N - - - - - Q Q O P V I T A K T A H Q H Q H M Q O F M H Q M Q I Q R Q H M Q - - - - - 3849
2168	Qy	S H V P S E A Q P T H A O S S K P V A A Q S O P S N V G Q S P V R V Q S P S Q T R I P S T P S Q L S P Q Q - Q S 2226
3850	Db	- - - - - Q Q O L H G S Q - - - - - Q I T S A P Q H M Q H Q Q A Q Q Q Q H N Q O H L N Q Q L H A Q O H P T Q K O H 3901
2227	Qy	Q V O T T T S O P T P I O P H T S L O I P S Q O Q S O P O V Q S - - - - - S T Q T L S S G T T L N Q V S V S S P R P Q 2283
3902	Db	Q A O Q P N Q - - Q I O Q H S Q Q - - - - - Q H V Q Q N Q A Q Q H L S Q Q Q H S Q Q O L N Q - - - - - Q H A Q Q 3952
2284	Qy	L Q I Q Q P Q P - Q V I A V P Q L O Q Q - - - - - V Q V L S O I Q S O V V A Q I O A Q S G V P P Q Q I K 2329
3953	Db	Q Q L Q Q I Q K L Q M H G P Q O Q K S P Q G V G H L G S T S I F A S Q H N S O L P A R - - - - - G V P Q Q - - - - - 4004
2330	Qy	L Q I P I Q I Q S S A V Q T H Q I Q N V - V T V Q A A S V O E Q L O R V - - - - - Q O L E D O Q K K Q Q Q O I E K R H 2386
4005	Db	- Q H P Q O L S H S C K P N T L V S N V G O V P P A I - - - - - L T R V G S H S Q N Q O Q Q L P H Q Q S S G H P H 4060
2387	Qy	T L Q A S N O S E T I Q R Q - - - - - V V M K H N A V I - - - - - E H L Q K K S M T P A E R E E N Q R M T C V N Q V M K Y I L 2440
4061	Db	Q K L S S P G A N L P L O T P L N V I Q N T P K I I V Q Q H I A V A Q N V P P Q T O G N A I H V P Q N Q - C K D S T 4119
2441	Qy	D K I D K E Q K A A K R K R E E S V E Q R K S Q N A T K L S A L - - - - - - L F K H E Q L R A E I L K R A L 2493
4120	Db	P P G H V E P T P A M S A Q K T S E S V S I R T P T T P T G L A V I S A N T V G S L T T E N L I K I S O P K O D E L 4179
2494	Qy	L D K D L Q L E V Q - - - - - E E L K R D L K I K K - - - - - E K D L M Q L A Q A T A V A A P P P V T P V L P A P 2541
4180	Db	I E Q S K - E V D S D T W S A K E V N I D S V I K L D T P L A S K A K R A V E M Q A I - - - - - - - - - - - - - - - 4224
2542	Qy	P A P P S P P P P G V O H T G L L S T P T L P V A S O K R K R E E K D S S K S K K K M 2589
4225	Db	- A P A I P N P O R G N O - - S M A O E T A L P T T S M S V N N S N D H D T E D E T R O L 4269

Search completed: September 24, 2003, 01:11:28
Job time : 64.7816 secs

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	3363	23.5	810	2	G01252	small GTP binding	
2	871.5	6.1	1711	2	T21432	hypothetical prote	
3	522	3.6	2447	2	T16870	hypothetical prote	
4	494.5	3.5	5327	2	T13564	microtubule-associ	
5	459	3.2	5170	2	T15348	hypothetical prote	
6	427.5	3.0	3759	2	A35085	trithorax protein	
7	419.5	2.9	2232	2	T34434	hypothetical prote	
8	413.5	2.9	3507	2	T34513	hypothetical prote	
9	412	2.9	2897	2	B48666	cell proliferation	
10	404.5	2.8	452	2	T21435	hypothetical prote	
11	404.5	2.8	3256	2	A48666	cell proliferation	
12	404.5	2.8	3488	2	T34418	hypothetical prote	
13	402	2.8	1630	2	A53077	ascites sialoglyco	
14	401.5	2.8	1367	1	A8478	glucan 1,4-alpha-g	
15	399.5	2.8	405	2	T21433	hypothetical prote	
16	399.5	2.8	510	2	D11430	hypothetical prote	
17	399.5	2.8	2441	2	T71623	erythrocyte membra	
18	397	2.8	2346	2	T13829	tpx homolog - frui	
19	395.5	2.8	2271	2	F90073	hypothetical prote	
20	394	2.7	2938	2	T30249	cell proliferation	
21	387.5	2.7	1229	2	T25697	hypothetical prote	
22	382.5	2.7	1589	2	T13606	hypothetical prote	
23	382	2.7	2453	2	S60254	nuclear receptor c	
24	375.5	2.6	3924	2	S37431	ankyrin 2, neurona	
25	374.5	2.6	2116	2	A26655	myosin heavy chain	
26	374	2.6	4377	2	A55575	ankyrin 3, long sp	
27	373.5	2.6	2541	2	T29340	hypothetical prote	
28	372	2.6	5105	2	T32650	hypothetical prote	
29	367.5	2.6	3259	1	A56539	giantin - human	

481 SE-----482
481 SEPTVEGDKGNSVANLGDNTTATSETPSEGRSPVGCSETPDSSNMAEKKVASL 540
483 -----482
541 PDVPEPNTCESNTSATTTSTQPNLENSNSSLSSQESAKAADPENGERSHT 600
483 -----VGFSEKSENGELSESPGAGKASGSTRITIRLPDPSKLSQKSOQVAAAA 534
601 PVSIOEIVGDTSEKSTGELSESPGAGKASGSTRITIRLPDPSKLSQKSOQVAAAA 660
535 HEANKLKEGKEVLVNSQGEISRLSTKKEVIMKGNLNNYFKLGOEGKYRVYHNOYSTNS 594
661 HEANKLKEGKEVLVNSQGEISRLSTKKEVIMKGNLNNYFKLGOEGKYRVYHNOYSTNS 720
595 FALNKHQHRDHRRHLAHLKFCITPAGEFKWNSVHGSKVLTITSLRLTITOLENNIPS 654
721 FALNKHQHRDHRRHLAHLKFCITPAGEFKWNSVHGSKVLTITSLRLTITOLENNIPS 776
655 SFLHPNW 661
777 TSLHPSF 783

RESULT 2
T21432
hypothetical protein F26H11.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21432
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21432
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1711 <WIL>
A:Cross-references: EMBL:Z81515; PIDN:CAB04197.1; GSPDB:GN00020; CESP:F26H11.2
A:Experimental source: Clone F26H11
C:Genetics:
A:Gene: CESP:F26H11.2
A:Map position: 2
A:Introns: 56/3; 154/2; 373/2; 417/3; 505/2; 596/1; 703/2; 847/3; 922/3; 1034/2; 1078/2;

Query Match 6.1%; Score 871.5; DB 2; Length 1711;
Best Local Similarity 17.7%; Pred No. 5, 3e-25;
Matches 420; Conservative 299; Mismatches 618; Indels 1041; Gaps 71;

4 EEEEE---DGAETQDSEDE---EDMEEDDDSDYPEEMEDD---DDDASYCTE 52
169 KRQEDIIYMDSEEEESDDFMLENDQVQVEEELNLTIDIKIEKGLDEENKYC-- 226
53 SSFSSHSTYSSTPGRRKPRVHRPRSPILIE--KDTPLPEFKSSDLAVPNEHIMNVAI 110
227 -----PWLEDDPASLPKLELPSSQDIPITPASIMDAVEI 261
111 YEVLNFGTVLRISPRFEDFCAALVSOQCITLMAEMHVVLKAVLREEDTSNTTFFPAD 170
262 YEILRSYHRTLRTPTTFEDFCAALISHNNSCIMAEVHMLNRLKSDDEBOTHYSVTE 321
171 LKDSVNSTLYFDIGMTWPEVLRYVYCESKEYHHVLPYQEA-----EDPYGPVENKI 222
322 TNSNVNIMHHMDTLTYABILRQYIEA-----YPFADASVRDAINDVNDVPFGYDANI 374
223 K-----VLQFLVDQFLTNTNIAEELMSGVIOYDDHCRVCHK-LGDLIC 265
375 QRFDSFFFNKHFVRLVLLFMSYRFLYSSEFKLVNVNGKFGQNDENCRVCGKSSRVVG 434
266 CETCSAVYHLECVPKPLEEVEDEMOCEVCAH-K-VPGVTDCAVEIQKNNKPIRHEPIGY 324
435 CTCOEAAHFVEC--SHLKPFPE-VLCNICKKNSAVRGVLPDPEAVDREP--LRSQPIGR 489

QY 325 DRSRRKYWFLNRRLLIEEDTENENEKKIWIYSTKVQLAEILDCLDKDYWEAECLKILEM 384
DB 490 DRYGRYWFIVRLVQVSLDETE---LYYSTVQVQLQKLDRTYYEKLCDTIRL 545
QY 385 REEIHMDITDLTNKAGSNKSFLLAANBEILSIRAKK---GDIDNVKSPE-ETEKD 440
DB 546 IDEFLQMALTVEMTSERR-----EAALETWAKRQLIGDYFAEATTPOYLHLD 594
QY 441 K-NETENDSKDAEKNEFEFQDSEKSDDKTP-----DDPEQCKSVGDFKSEKNSG 493
DB 595 SKRMASIIURDCAQKQVQKVEEPVQSPVKCVQFVEDSILLPESMIGIF----- 647
QY 494 ELSPEAGKAGSGASTRIITIRLNPDSKLSQKSOQVAAAAHEANKLFEGKEVLVNSQ 553
DB 648 -----DAKL-----INTFWSGG----- 659
QY 554 GETSRLSTKKEVI-----MKGNNIN---NYFKLGQSGK---YRVYHNOYSTNSFALNKHQ 601
DB 660 -----ATOEELEYEQVDISDNFDAPSANLWRMGDEGNDQTFWTYTYNYSRNEMSESFLT 713
QY 602 HREDHDKRRHLAHLKFCITPAGEFKW---NGSVHGSKVLTITSLRLTITOLENNIPSSFL 657
DB 714 RKKAADKKYMASKF--AQIDNFDWVVAKNRQFYGDASLHCKFIMWTLOQVINKPIDLM 771
QY 658 HPNWSHRANWIKAVQMSKPRFALALALECAVPVWMLPIWRFLGHTLRLHRTSIE 717
DB 772 HRKWEFAKGFDELSVADDDYKKLVTCLLKLDCAVRKTIIFMPQMWNGLGOTRLERITVQ 831
QY 718 REEKEKVKKEKK-----QEEETMOQATWYKTYFPVKHQWVKQKGEYRVTG---YG 767
DB 832 RENFMKQQRLKKIDADALTKDLDSDSFVRVYVNMKPKWPTY--ILRQGETYRNAGKSGM 890
QY 768 GWSWISKTHYRVFVKPLPGNTNVNRYKSLEGTNNMDENDESKKCSRPKKIKIEPD 827
DB 891 GWAWVAKYVEKWI-----QVPESKPLPLAVTVEIEIKTESV 926
QY 828 SEKDEKVGSDAAKAGADQENMDISKITEKDDVKELLDSDSDRCKPEEPMEVDDDMKTES 887
DB 927 SNRK-----ARLELLVSKIT----- 942
QY 888 HVCQESSQDVVNVSEGFHLRTSYKKTKTKSKLDGLLERRIKOFTLEEKORLEKIKLEG 947
DB 943 -----KKRQSG----- 949
QY 948 GIKGIGTSTNSKNLSESPVITKAKEGCQSDSMRQEQSPNANDQEPDLIOGCSQSDSS 1007
DB 950 -----GKSSKKPTFELT-----NGCYSPSCRS----- 971
QY 1008 VLRMSDPHTTNKLYPKDRVLDVSIKSPETKCPKQNSIENDIEEKVDSLARGQEPKTS 1067
DB 972 -----NPNRKC-----YSP--- 980
QY 1068 KTKGNDFFIDDSKLASADDITGLTICKNNKPLIQEESDTIVSSSKSALHSSVPKSTNDRDA 1127
DB 981 -----MCRN----- 984
QY 1128 TPLSRAMDFEGLCGDSESNTLENSDTSVIQDSSSEDMIVQNSNESISEOFRTREQDV 1187
DB 985 ----- 984
QY 1188 EVLEPLKCELVSGESTGNCEDRLPVKGTANGKKPSQOKKLEERPVKNKSDQIKLKNWTD 1247
DB 995 -----GYLVSQAKQAHDERKLEE----- 1001
QY 1248 KKNENRESEKKGORTSTFQINGKDNKPKYILKGECLKEISESRVSVGNVPEKVNINKI 1307
DB 1002 ----- 1001
QY 1308 IPENDIKSLTVKESAIRPFINGDVIMEDFNERNSETKSHLLSSDAEGNYRDSLETIPS 1367
DB 1002 ----- 1001

QY 1368 TKESDSTQTTTPSASCPSNSVNOVEDMEIETSEVKKVTSPTSEESNLSNDFIDENG 1427
Db 1002 -----
QY 1428 LPINKNENNGESKRKTVITEVTTMTSTVATESKTVIKVEGDKQTVVSSSTENCAKSVT 1487
Db 1002 -----SCVLGEK-----
QY 1488 TTTTTVTKLSTPSTGGSDIISVKEQSKTVVTTVTDLSLTTTGGTLTMTSVKEYSTRD 1547
Db 1010 -----
QY 1548 KVKLMKPSRPKTRSGTALPSYKRVTKSKSIFVLPNDLKLARKGGIREV--PVFN 1605
Db 1010 -----AMPIDQTFSTKRGKGSIFVLQKTLRQIMGGGQQQYMPGFS 1054
QY 1606 YNAKPALDIWPPRPRTFGTTWRYLQTVKSLAGVSLMLRLWASLRWDDMAAKVPPGG 1665
Db 1055 AGIKSNLLIWPYPAPRPTLDLCWKQTLNARSLHVALQLIWSIKFNEFD---PDPT 1111
QY 1666 GSTRTTSET--EITTEIIKRRDVPYIGIRFYCIKRIICPI-GVPETPKETPTQRK 1722
Db 1112 HPDRRVVIDTSHDERRRIIRKEMPPYQYVERVEMEIEIIPLYDEPEEDESNLRNG 1171
QY 1723 -----LRSSALRPKRPET-----PKQT-----GPVIIETWVAEELE 1754
Db 1172 GSSEFSSRRSSARKRQRHEFLSKFGNYPKKNAFRSLDNRRAIRREWDDGVTLK 1231
QY 1755 LWE-----TRAFARVEKEKAQV-----EQAKKRLQOKQPTVIATSTTSPTSST 1800
Db 1232 VFEIKDWYKWRARAEAKTAKKLEATRAKAQAKAEDEERRIQOQQOQSVARIPV-PMHS- 1289
QY 1801 TSTISPAQKVMVAPISGV---TTGKMWLTITKVGSPATVTFQON-----KNFHQTF 1849
Db 1290 ---LIPSENNV-PYLGSOQQRRPNNGERNERGFLEKYNSSSVSPQAHGVASTPPPGYHQ 1345
QY 1850 ATWKQGSNGVVOQKVLGIIPSTSTGTSTQOFTTFQPTATVTRPNTSGSGGTTSN 1909
Db 1346 PNIIRQAGYNQ-----LPRKPTTSPFNFSQ-RP-VATIPTTQPLRAAGADG- 1390
QY 1910 SOVITGQIRPG---MTVIRTPLOQSTLGAIRTPVMVOPGAPQVMTQIIRGQ--- 1961
Db 1391 --VVRVNMVTPCNKSTVNTSTPYQ-----ALNRQYQLORQOQQFAVRLTNGYHFMD 1443
QY 1962 -----PVSTAVSAPNTSVSTPGQKSLTSAT----- 1986
Db 1444 GTMRGGGRNPSVMQHRLPQNRALQRPFGESTTEMRVRTEAIPDNDGDDEQPPVIPRYD 1503
QY 1987 STSNIOSSASQPPRPOGOVKLTMAQLTQLTQGHGGNGLTVVTOGOGQTTGQLQL- 2042
Db 1504 PTSNFDQRAQOQHPQSRPVYSTPAQMIRTTPQGVKH--NVILMKASDGTQKRVLPKG 1561
QY 2043 IPQGVTVLPQGOQLMAAMPNGTVQRFLEPLATTATTATTTTSTTAAGTGEOROS 2102
Db 1562 FPPGTVI--SIGQRVVYRQPTAVQOQLYT-----ATPGTRVVRIPNANGAPRQODH 1613
QY 2103 KL-----SPQMVGHQDK--TLPPAQ-----SSSYGPAKAQOQPTAQPSARP 2140
Db 1614 QVMRRVVYQASPRAMEYMDQGTTPPGQOQVRYVLOGGNSGTPNPPKVSRRGGPRGLT 1673
QY 2141 QPQTPQSPAQPEVQTOPEVOT---QPTVSHVSEBAQ 2175
Db 1674 MQMVQOQOQHNPAPHYDPPDATGFAVSTTEQVPEQ 1711

RESULT 3

T16870

hypoetical protein T13H2.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Sep-2000

C:Accession: T16870

R:Wu, X

submitted to the EMBL Data Library, October 1995

A:Description: The sequence of C. elegans cosmid T13H2.

A:Reference number: Z18593

A:Accession: T16870

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2447 <WUX>

A:Cross-references: EMBL:U39653; NID:gl049397; PID:gl049400; PIDN:AAB52494.1; GSPDB:G

A:Experimental source: strain Bristol N2; clone T13H2

C:Genetics:

A:Gene: CESP:T13H2.3

A:Map position: X

A:Introns: 18/3; 135/3; 176/3; 496/1; 693/1; 816/1; 1053/1; 1131/3; 1181/3; 1308/3; 1

C:Superfamily: RING finger homology

F:158-207/Domain: RING finger homology <RRN>

Query Match 3.6%; Score 522; DB 2; Length 2447;

Best Local Similarity 18.5%; Pred No. 9e-12;

Matches 502; Conservative 413; Mismatches 1033; Indels 760; Gaps 114;

QY 9 EDGDAEFTQDSEDEDEDEDDDDSDYPEEMEDDDDDASYCTESS----- 54

Db 2 DSDSPGSTSKSARDKAENAEENTSDSSDSEVSSASEKSESRPSEKKKVVITRIVPVRP 61

QY 55 -----FRSHSTYSTPGRRKPRVHRPRSPILPEKDIPLEFPKSSDLMPNHEIMVVI 108

Db 62 PTRDKGHRVNLLESGNESETKSLYQRAKEGIPSYKGPKEIKLPTTSEQYDLEEVLMN-- 119

QY 109 AIYEVLRNFGTVLRSLSPREFDCAALVSQCTLMAEMHVLLKAVLREEDTSNTTGP 168

Db 120 -----PARMEG-----RELTLNA----- 132

QY 169 ADLKDSVNSTLYFDIGMTWPEVLRVYCESDREYHHVLPYQAEVYPYGVENKIKVLQFL 228

Db 133 ---YDAVRNKYNLPG-----KSVCEAD-----LQKV 156

QY 229 VDQFLTNIAREELMSEGVIOYDDH--CRVCHKLGDLCCETCSAVYHLECVKPPLEVP 286

Db 157 IGSF-SCDVCQELIQGSIMTKCGHRFCDCILVAFMRSGNTCTPTCRNLGSKRELOQDP 215

QY 287 E-DWQCQEVCAHKVPGVTDCAVEIQKN-----KPYIRHEPIGYDRSRKYWFLENRR 337

Db 216 RFDQLLIQVVSRSIVG--RMAENREHEKDVYFGKGYIEG--GSDWNRKYGDIPNSK 270

QY 338 LIIEDTENENEKKI-WYYSTKVQLAELIDCLDWDYEAELCKLILEEMREEIHRHMDITE 396

Db 271 LKAPRLKSAKGRKKIRWFHES-----DED---GSVRKVMESKKA-----PKE 310

QY 397 DLTNKARCSNKSFLAANEEIILESTRAKKGDID---NVKSPETEKKDNENETNSDKAEK 453

Db 311 DDTNYLENDKEGTSVAAEKEVLE-----EGEMDFPIEIKSDDEQTDLDDEESLSDSF 365

QY 454 NREEFQDQSL-----EKDSDDKTPDDPQEQGSEVGDGFKSEKSNSELS 498

Db 366 EISDNEDVSKSCSTSKTTNRSRSDSSDNDSDRNLQKKRKKRKNVPTDGDSDVSN 425

QY 499 PGACKGASG-----STRIITRLR-----NPDSKLS-OLKSQQVAAAAHAANKLFKEGKVLV 549

Db 426 ESFEDASGVEVATKLKESKKKCGRPKKFAPELIEGDIPTPSDLSITSSDERDDNA 485

QY 550 VNS-----QGEISRLSTKKEVIMKGNINNY-----FKLQGEKRYRYHNYQSTN 593

Db 486 ADPYAFVQKEFNRDPRRDGHPEKDKLYNFDMDNMHVDRKPEKGEIHVISD--DSN 543

QY 594 SFALNKHQHRDHDKRRHLAHKFCITPAGEFKWNGSVHGVSKVLITSLTLTLTQLENNIP 653

Db 544 S-----EHESDEAEDRESSI-----DSEHEKEI----- 566

QY 654 SSFLHPNASHRANWIKAVQCMCKSPREFALALALECAVKPVVMLPIWREFL-----GHTR 709

Db 567 SKFL-----SHR-----QPLPNPTS-----VDDCCQVITVVKKDKVQSAITSKPGETS 609

QY 710 LHRMTSITEREKEKVKKKEKQEEETMQQATWTKYTFPVKHQVWKGGEYRTVYCGW 769

A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1

C;Genetics:

A;Cross-references: FlyBase:FBgn0025392

A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1

A;Note: EG:4954.1

C;Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 3.5%; Score 494.5; DB 2; Length 5327;

Best Local Similarity 17.7%; Pred. No. 2.6e+10;

Matches 527; Conservative 478; Mismatches 1189; Indels 791; Gaps 119;

QY	1	MVSEEEEDGDAEETQD--SEDDDEDEMEEDDDSDYDEE-----MEDDDDDASYC 50	2512	ISR--PASAGETASSPIEAPKDFAEFEQAEKAVLPLTIELGNLPTLSSPVDVAHASVQ 2569
DB	1751	LASEASRPASVAESVKDAEAKESKEESRRESVAEKSPLPSKESRSPASVAESIKDAEKS 1810	805	-ENMDESDRKCRS-----PKKIKIEPDSEKDE-----VKGSDAK----- 840
QY	51	TESSFRSHSYSTPGRRPRVHRPRSPILEEKDIPPLEFPKSSSEDLM-----VPN 101	2570	PAELSKVDTIEKTASSPIDEAPKSLIGSPAEREERESPAESAADAESVEKSKDASRPSPV 2629
DB	1811	KEES--RRESVAEKSPLPSK--EASRPASVAESIKD---EAEKSKEESRRESVAEKSPLPS 1864	841	---GADONEMDIS---KITEKKDDQVKELLOS-----DSDKPCKEPEMEVDD 881
QY	102	EHIIMVIAIYEVLRNFGTVLRSPFRFEDFCALVSOEQCTLMAEMHVLLKAVLREEDT 161	2630	ESTKADSTKGDISPSPESVLEGGPKDDVEKSKESRSPSPASITGDSTKDV--SRPASVVE 2688
DB	1865	KEASRPASVAESIKD-----EAEKSKEE-----SRRESVAEKSPLPSKEA 1904	882	DMKTESHVNCOSSOVVNVNVEGFHLRTSYKKTKKSSKLDGLLERRIKOFTLEEKORLE 941
QY	162	SNTTFGPADLKDSVNSTLYFIDGMTWPEVLRVYCESDKYHHVLPYQEAEDYPYGPVENK 221	2689	SVKDEH--DKAESRRESIAKVES--VIDEAGKSDSKSSSQD-----SOKDEKSTLASKEASR 2741
DB	1905	SR----PASVAESIKDE-----EAEKSKEESRRESVAEKSPLPSKESRSP 1944	942	KIKLEGIGIKIGITSTNSSKNL-----SESPVTK--AKEGCQSDSMRQEQSPNA 989
QY	222	IKVLOFLVQFLTNTIAREBELMSEGVIOYDDHCRVCHKLGLDLCCTCSAVYHLECVKP 281	2742	RESVVESSKDDAEKSERSPESVVIASGEPVPRESKSPLDKDTSRPGSVSVESVTAEDKSE 2801
DB	1945	ASVAESIKDE---EAEKSKEESRRESVAE-----KSP 1972	990	NNQPEDL---IOGCQSOSDSSVLRMSPDPSHTTNKLYPKDRVLDVDSIRSIPETKPKONSI 1046
QY	282	LEVPDENOCEVCVAHKVPGVTDVCAETQKNKPYIRHEPIGYDRSRRKYFWLNRLLIIE 341	2802	QOSRRESVAESVKADTKDKGKQEAERPSVDELL--KDDDEKQESRRROSITGSHKAMST 2859
DB	1973	L---PSKEASRPASVAESIKD-----EAEKSKEESRRESVAE--EKSPLPSKEASRPASVA 2022	1047	ENDIEEKVSDLASRGOEPTK-----SKTKGNDFFIDDSKLASADDIGTLICKNKK 1096
QY	342	EDTENENEKKIWTYSFKVOLAEILDCLDKYWEALCKLLEEMREIHRHMDITEDLTNK 401	2860	MGD--ESPMDKADKSKERPSRPSVAESIKHENTKDEESPLGSRDRSVAESIKSDITKGEK 2917
DB	2023	ESIKDAEKS--KEESRRESVAEKSPLPSKE--ASRPASVAESIKDAEK-----SKEESRR 2075	1097	-PLIOQE---SDTIVSSSK-----SALHSSVPKSTNDRDATPLSRAMDF--EGKLGC 1142
QY	402	ARGSNKSFTL---AANERILESIRAKGIDIDNVKSPETEKKDNETENDS--KDAEKNRE 456	2918	SPLPSKEVSRPESVVGSIKDEAKESRRESVAESVKPESSKDATSAPPSKEHSPESVLGS 2977
DB	2076	ESAAEKSPLPSKEASRPASVAESVK-----DEADKSKEESRRESMAESGKAOS 2123	1143	DGESNTLNSSTVSIQDSSSEED---MIVQNSNESISEQFRTREODVEVLEPLKGCELVS 1199
QY	457	EFEDQS-----LEKSDDDTPDDDPQKGSEVDFKSEKNGELSSPGAGKA 505	2978	LKDEGD--KTTSRVSVADSIKDEKSLVSOEASRPESE-----AESLKDAAP 3024
DB	2124	IKGDSPLEKVRSPESVAESVKKDDPVKSEPSRRESVAGSVTADSARD--QSPLESKGA 2181	1200	GESTGNCEDRLPVKGTENGKKPSQOKKLEERPVNKCSDQIKLKNNTDKNNNRSEKK 1259
QY	506	SGSTRIITRLRNPDSKLSQLKSOQVAAAHAANKLFKEGKCVL-----V 549	3025	SOETSRPES---VTESVKDGKSPVASKE--ASRPASVAEN---AKDSADESKEORPELPO 3077
DB	2182	SRPESVVDVSKDEAEKQESRRRESKTESVTPPKAKDDKSPKEVLQPVSWTFETIREDAQPM 2241	1260	GORTSTFOINGDKNKPIYLGKCLKEISESRVVGNEVPKVNINIKIIPENDIKSLITVK 1319
QY	550	VNSQGEISRLS-----TKKEVIMKGNINNYFKLQGEKGYRYVHNOYST 592	3078	SRAGSI---KDEKSPASKDAEAKESRRES-----VAEQFP-----LVS 3116
DB	2242	KPSQAEERRESIAESIKASSPRDEKSPLASKEASRPGSVAESIK-----YDL 2288	1320	ESAIRPFIINGVIMEDFNERNSETKSHLLSS-----SDAEGNYRDSLETLPETKESDST 1374
QY	593	NEFALNKHQHRDHDKRRHLAHKFCFLTAPAGEPKWNGSVHGSKVLITSTLRLLITOLENNI 652	3117	KEYSRPASVAESVKD---EAEKSKEESPLMSKEASRPASVAGSVKDEAEKSKEESRRESV 3173
DB	2289	DKPQIHKDKSTEHRSRESLEDKSAVTS-----EKSV--SRPLSVASDHEAAVAATEDDA 2340	1375	QITTPSACSPENSVMQVEDMEIETSEVKK-----VTSSPITSEESNLSNDFIDENG 1427
QY	653	PSSEF-----LHPNW-----ASHRANNIKAVQMCSS----- 676	3174	AEKSPULPSKEASRPASVAESVDEADKESRRESGAEKSPLASKEASR----- 3223
DB	2341	KSISPKDKSRPGFVAETVSSPIEATMEFSEKIEVVEKSSLSLALSGGGGKLQTDSSPV 2400	1428	LPINKNENYNGESK-----RKTVITEVTMTSTVATESKTVIKVEKGDKQTVVSSSTENC 1481
QY	677	--KPRFALALALECAVRPVVMLPIWREFLGHTR-----LHRMTSI 716	3224	-PASVAESIKDEAEKSKEESRRESVAEKSPLPSKEASRPTSVAKSVKDEAE---KKEES 3279
DB	2401	DVAEGDFSHAVASVS--TVPTLTLPKPAELAQIAAKTVSSPLDEALRTPAPAHISRADSP 2459	1482	AKSTVTTTNTTIVTK--LSTPSTGGSDVLIISVKEGSKTVVTTTVDLSLTTTGGTLVTSMTVS 1540
QY	717	EREEKEVKKKKQEEBETMQOATWVKYTFPVKQHVWKQKGEYRVTVGYGWSWISKTH 776	3280	SRDSVAEKSPLASKEASRPASVAE---SVQDEAEK-----S 3312
DB	2460	ABCASEETASQDKSPQVLKESRPAWVAESKDDAQL--KSSVEDLR-----SPVASTE 2511	1541	KEYSTPDKV--KLMKFSRPKKTSGTALPSYRKFEVTKTKKSIFFVLPNDLLKLARKGGIR 1599
QY	777	VYRFVPKLPDNT-----NWNYSKSLGKTKNMD----- 804	3313	KEESRRESVAEKSPLAYKEASRPASVAESIKDAEKSKEES-----RRESVAE 3360

1600 EYPYFYNNAKPAIDINWPPSPRPTFGITWRYRLQTVKSLAGVSLMLRLLWASLRWDDMAA 1659

3361 KSPLASKEAS-----RPT--SVAESVKDEAEKSKEESR-----DSVAE 3397

1660 KVPGGGSTRTETSETTEITTEILIKRRDVG---PYGIRFEYCIRKTIICPIGVPEPKETP 1716

3398 KSPLASKEASRPASVAESVQDEAEKSKEESRRESVAEKSPLASKEASRPASVAESVKDDA 3457

1717 TPQRKGLRSSALRPPRPETPKQTG--PVIIETWVAEELELWEIRAEARVEKEKAQAVEQ 1775

3458 EKSKEESRRESVAEKSPLASKEASRPASV-----AESV--KDEAEKSKE 3499

A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GEI>
A:Cross-references: EMBL:U08046; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Query Match 2.98; Score 419.5; DB 2; Length 2232;
Best Local Similarity 21.88; Pred. No. 5.3e-08;
Matches 339; Conservative 172; Mismatches 633; Indels 413; Gaps 63;

QY 945 LEGGIKIGITSTSSKLNSESPIITRAKEGQSDSMRQEQ-----SPNANDQPDLLQGC 1001
DB 731 LESTSG-ATYSSGAGTTWSP-----SQSSVSGSGGTSPPASTSGEMTSQGS 781
QY 1002 SQSDSSVLRMSDPSTTKLYPKDRVLDVYSIRSPETKCPKQNSIENDIEKVSDLASRG 1061
DB 782 TQTPGSSVSTSAALITSTQ-----QSVSTNSPGSTVTRPST-----VSGSTSSG 825
QY 1062 QEPKSKITKGNDFIDDSKLASADDIGTLICKNNKPLIOESDTIV-----SSSKSLHS 1116
DB 826 STVTGSTEAS---TSGSSVASSSPAPS-TSQNPSTSGSSMITQSPYPSOSTSPVES 881
QY 1117 SVPKSTNDROATPLSRAMDFEGKLGCDSESNSTLENS---SDTVSIQDSSEEDMIVQNSN 1173
DB 882 STTSPGSGPTTLTSTS-----PSPQSITIGTQGSTPGISTTSEE-MTSGST 931
QY 1174 ESISEQFTRBQDVEVLEPLKCELVGSESTGNCEDRLPVKRGTEANGKKPQQKKLERPV 1233
DB 932 QTPGSTGSTVTPQSTVSD-----STSSGS---TVTVGSTECS-SSPIPSTQNTNPS 979
QY 1234 NKCSQJIKLKNITDKNNENRESEKKQRTSTFOINGKDNKPKIYLKGCLEKISESRV 1293
DB 980 TSGSSMSTQPPQSQSTSPVESSTSGATSSS-----GSP-----GTLTISPSPPSP 1027
QY 1294 S-----GNVEPKVNNIKIIP--NDIKSLTVKESAIRPFINGVIMEDFNERNSET 1344
DB 1028 SSTIGSSQGSTSPVVISQGSTETPGSTGTSTVTKPST---VSGSASSGSTATMGSTEA 1083
QY 1345 KSHLLSSDABGNVRDSELETLPSTK---EDSTQTTTPSACPSNSVNOVEDMEITSE 1401
DB 1084 SS---TSGGSSPNPQSSTSPSTSGATSPGSSGTLTISPSPSQSSITGSSQGST 1140
QY 1402 VKKVTSPITSEESNLNDPENGGLPINKNENVG-----ESKRKTVIT 1447
DB 1141 VVSTTSGDMSQGSTQIPG---STGSTVTPQSTGSGSTSTSGEITSGGSTQTPRSSLS 1196
QY 1448 EVTTMTST---VATES--KTIVKVEKGDQTVVSSSTENCAKSTVTTTTT-----TV 1493
DB 1197 SPAISTSTQOQSVSTNSPGSVTQ-----PSVIRGSTS--SGSTVTGSGSTSGSSA 1249
QY 1494 TKLST-----PSTGGSDVILISVKEQSTVTVTTVTDLSL-----TTGGT 1532
DB 1250 TSLSSSSPVPSTQSPNPSTSGSSTPTPNPQSSTSPVVTGEMTSHGSTQTPSTIGST 1309
QY 1533 LVTSMYTSKEYSTRDKYKLMKFSRPKKTGTALPSYRKFKVTKSTKKSIFVLNPDLLKL 1592
DB 1310 VTQSTVSGSSSGSTVTI---GSEASTSGSSF-----KTPSSISPVPTSS----- 1354
QY 1593 ARKGGIREVPYFNNAKPALDINWPYPPRPTF-GITWRYRLQTVKSLAGVSLMLRLWAS 1651
DB 1355 -----PIPSITTFASSTSGSTISDVSSVSTTSLA----- 1382
QY 1652 LRWDDMAKVPGGGSTRTETSETTEITTEILKRDVGPYGIREFYCIKIKIICIGVPET 1711
DB 1383 -----PLSSSLP-----STVPSSQTQFSSTSEGSRASSP-----VPSQ 1417
QY 1712 PKETPTQKGLRSALRPK-----RPETPKQGPVLIETWAE 1752

DB 1418 TSSTPTNPCTGSESSILLSTISGTOHTMTSKASSGSTSPSTNSQGTGTVTMG----- 1471
QY 1753 LELWEIRAFARVEKEKAQAEQQAQKRLBQKPTVIATSTTPTSTTSTTISPAQKVMV 1812
DB 1472 -----SSSTSGSVSTSSASTQPMSTSQGSSAGSTVASSTASPAASSTAPSSTG--TMS 1523
QY 1813 APISGSVTTGTTKMWLTTKVGS---ATVTFQONKNFHOTFATVWKQCSNSGVVQVQKV 1869
DB 1524 STSGTVGVTISESSTTASASSQTGTVTMGSSST-----SGVSTSSASSSTQPM 1573
QY 1870 LGIIPSTGTSQOTFTSFQPTATVTRPNTSGSGGTTNSQVITGPQIRPGMTVIRTPL 1929
DB 1574 STQGSAGSTVASSSTAGLVSTIV---PSTGTMGSTSG-----TVGSTIS 1618
QY 1930 QQSTLGRKAIIRTPVMVQPGAPQVMTQIIRGQPVSTAVSNPTVSVSPGQ-----KSL 1982
DB 1619 ESSTTASASSQTGTVTMGSS-----TSGVSTSSASTQPMSTSQGSSAGSTVASST 1672
QY 1983 TSATSTSNIOSSASQPPRPOGQVKLTMAOLTQGHGNGOGLTVVIQGGQT----- 2036
DB 1673 TGLVSTSTVPSSTGTMGTSSTGTVGVTISESS--TAASASSQGTGTVTMGSSSTGVS 1730
QY 2037 ---TGQLQLIPQ-----GVTVLPGPGQOLMOAAMP-----NGTVQRFLEFTPIATT 2078
DB 1731 SASGQQPMSTSQGSSAGSTVSVSTASPAASSTAPSSTGTMTSSTSGTVG-----STMSQS 1786
QY 2079 ATTAATTTTTTAACTGEGQKSLSPQMVHQDK-----2114
DB 1787 STAASTTSHGTSTVTLGS-----SSTSNQMTSGSSVGVSTVASSTAGLVSTVPSSTG 1842
QY 2115 TLPPAQSSSVG-----PAKAQPTAQ-----PSARPOPTQPSAPQPEVQTQPEVQ 2161
DB 1843 TMGSTSGTVGVTISESSTTASASSQGTGTVTMGSSSTGVSSTSSASTQPMSTSQGSS 1902
QY 2162 TQTTVSSH-----VPSEAOPTAQS-----SKPQVAQSQPSNVQCSQSPVRV 2204
DB 1903 AGSTVASSTAGLVSTVSVSTGTMGTSSTGVTGSESTSESTAASTSSQITG---SIVTI 1958
QY 2205 QSPSOTRIRSTPQLSGQSOQVQTTSTQPIPIQHTSLQIPSGQGOQPOQVOST-Q 2263
DB 1959 GSTSGT--NPSSPRSL-----SQI-TITPSPQSSTESTQTLSPSSSPSTHVSSEGT 2010
QY 2264 TLSSGQTLN-----QVSVSPSRPQLIQOPOQVIAVPLQOQOVLSQ 2308
DB 2011 TMSGATTSGDKMSFLSTGTTVSFSSRGSLATTSAPKPSVTCFLMWTQSKIDQ 2067

RESULT 8
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favella, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/3504/1

Query Match 2.98; Score 413.5; DB 2; Length 3507;
Best Local Similarity 19.88; Pred. No. 1.6e-07;
Matches 263; Conservative 176; Mismatches 469; Indels 417; Gaps 48;

Db 415 HIALSENILGQFQCTGSEEPDLLPTSESFGNVFFSAQNAKQPSD-----KCS 466
QY 692 VKPVVMLPIWREFLIGHT-----RLHRMTSIERE-----EK 721
Db 467 ASP-----PLRQOCIRENGNVAKTPRNTYKWTSLSTKTSDETETBPSKTVSVNNSGRSTEF 522
QY 722 EKVKK--KEKQEEETM-----QOATWVKYTPPVKHQVWKQKGEERYVTGYGWS 770
Db 523 RNIOQLPVESSEETNIEVICILKRGOKATLLQO-----RREGEMKEI----- 566
QY 771 WISKTHYRVFVKLPUGNWNVYKSLGKTKNNMDENDESCK-----RKCS--RSPKKIK 823
Db 567 -----ERPFET--YKENIELKEN--DEKMKAMKRSWTGCKCAPMSDLTDLK 609
QY 824 IEPDSE--KDEVKG-----SDAKGADONEMDISKITEKKDQDVKELLSDSKPKCE-- 874
Db 610 SLPDTEMLKDTARGONLLOTDHAKAPKSEKGIITKM-----PCOSL 651
QY 875 EPMEDVDDMDKTESHVNCQESSQDVNVNSE-----GFHLRTS----- 911
Db 652 QP-----EPINTPTHTKQOLKASLGKVGVEELLAVGKFTRTSGETTHREPADGDKSIR 707
QY 912 -YKKTGK-----SKLDGLLERRIKQFTLEBKQKLEKIKLEGGIKGKGTSTNSKNLSE 965
Db 708 TFKESPKQIILDPAARVTGM--KKWPRTPKEEAQSLDL---AGFKELFOTPGPSEESMTD 762
QY 966 SPVITKAKGQCSQMRQEOPSNANNDQPE--DLIOGCSQSDSSVLRMSDPSTHTNKLYPK 1024
Db 763 EKT---TKIACKSPPPESVDPTSTKOWPKRSLRKADVEEFLALRLKLTSPSAGKAMLTPK 819
QY 1025 DRVLDDVSI-----RSPETKCPKONSIN-----DIEKVS 1055
Db 820 PAGGEXDKIAFMGTPVQKLDLACTLPGSKRQLOTPKEKAQLEDLAGFKELFOTPGHTE 879
QY 1056 DIASRGO-----EPKSTKGNDFIDDSKLASADDITGLI--CNKKKPLI 1099
Db 880 ELVAAGKTTKIPCDSPQSDPVDTPSTKQR-----PKRSIRKADVEGELLACRNLMP-- 931
QY 1100 QEESDTIVSSKSAHSVSPKSTNDRD-----ATPLSRAMDFEGKLGCDSESNTLENS 1154
Db 932 -----SAGKAMHTPKPSVGEKDIIPVGPVQKL-----DLTENLT 968
QY 1155 DTVSIQDSEEDMIVQNSNISISQFRTREQDVEVLEPLKCELVSGESTG--NCEDRLPVK 1213
Db 969 GSKRRPQTPKEEAQALEDLTGFKELFOTPGHTEEA-----VAAGTKTKMPCESSPPES 1021
QY 1214 GTEANGKPSQOKLEERPVNKCSDQIKLNTDKKNENRESEKKGORTSTFOINGKDN 1273
Db 1022 ADTPTSTRQKPTPLEKRDVOKELSALK-----KLTQSGETHHTDKVPGGED 1069
QY 1274 KPKIYLKGECLKEISESVVSGNV--EPKVNINIKIIPENDIKSLTVKESAIRPFGDVI 1332
Db 1070 KSINAFRETAQOKLDPAASVTGSKRHPKTK--ERAQPLEDLAGW--KELFQTP-----V 1119
QY 1333 MEDFNERNSETKSHLLSSDAENYRDSLETLPSTKESDSTQTTTSPASCPESNSNOV 1392
Db 1120 CTDKPTTHTEKTKIACRSQPD-----PVDTPTSSKPSQKRSIRK 1158
QY 1393 EDMETETSEVKKVTSS-----PITSEESNLSDNDFIDENGLPINK---NENVAGES 1440
Db 1159 VDVEEEFLALRRKTPSAGKAMHTPKPAVSGEKNIYA--FM---GTPVOKLDLTENLTGSK 1213
QY 1441 KRKTIVITVTTWTSIVA-----TESKTVIKVEKGDQKQTVVSSSTENCAKSTVTT 1488
Db 1214 RRLQTPKEKAQALEDLAGFKELFOTRGHTEESMT---NDKTAKVACKSSQPDLDKNPASS 1270
QY 1489 TTTTIVKLTSTGSGVDIISVQSKQTVTTVTVDLSLTGGLVMTVSVKEYSTRDK 1548
Db 1271 KRRLKTSLG--KVGVEKELLAVGLTQSGETHHTHTPTGDKGSKMAFMESPKQILDSA 1328
QY 1549 VKLMKFSRPKTRSGTA-----LPSYRK--FVTKSTKKSIFVLPNDL--- 1589

Db 1329 ASLTGSKRQRLTPKGSKEVPEDLAGFIELFQTPSHTKESMTNKTTKVSVRASQPDLDVT 1388
QY 1590 -----KKLARKGIREVPYFNYNNAKPALDIWPPSPRPRTFGITWVRLOTVKSLAGV 1641
Db 1389 PTKSKQPKESLKAATEE-----EFLAFRKQTPSAG---KAMHTPKPAVGE 1432
QY 1642 SLMLR--LLMASLRWDDMAAKVPPGGSGTRTETETETETETETETETETETETETETETET 1694
Db 1433 EKDINTFLGTPVQKLDQGNLP--GSNRRLOTRKEKAQALELT-----GFRELFTQ 1483
QY 1695 -----FEYCIRKILIC--PIGVP--ETPKETPTPQKGLRSS-----ALRKRK--- 1733
Db 1484 CTDNPTADEKTKKILICKSPQSDPADPTMTKORPKRSLKADVEEFLAFRLKLTSPAGK 1543
QY 1734 --ETPKQTPVILTIETWAEELWEIRAF---AERVE-----KEKAQA 1772
Db 1544 AMHTPKAA-----VGEK-----DINTFVGPVVEKLDLLGNLPGSKRRPQPKKAKA 1591
QY 1773 VEOAKKRLBQOQPTVIATSTTSPTSTTSTISPAOKVMVAPISGVTGTKMVLTKVG 1832
Db 1592 LEDLAGFKELFOTPGHTEESMTDDKITEVCSKSPQDPVKTP-----TSSKORLKISLG 1645
QY 1833 SPATVTFQKNKHQIFATWVKQGSNGVVQOKVL--GIIPSSTGTSTSOQTFSTFQPR 1890
Db 1646 K-----VGKVEVLVPGKLTQTSKGTQT----- 1669
QY 1891 TATVTIRPNTSGSGT-----TNSQVI-----TGQIRPGMTVIRTP-----LQOST 1933
Db 1670 -----HRETAGDKSIAKAFKESAKOMLDPANYGTGMRWP-----RTPKEEAQSLDLA 1718
QY 1934 LGKAIIRTPVMQPGAPQVMTQIIRIQPSTAVSAPNTVSSTPGQKSLTSATSTSIQS 1993
Db 1719 GFKELFOTPDHTEESTTDDKTKTI-----ACKSPPPESMDPTST----- 1758
QY 1994 SASOPPRPOGQGVKLT--MAOLTQTOCHGNGOGLTVVIOGOGTTQOLQIP-----QGV 2047
Db 1759 --RRRPTPLGKRDIIVEELSALKLTO-----THTDKVCPDEDKGI 1798
QY 2048 TVLPGQOQML-----QAAMPNGIVQ-----RFLF--TPLATTATTAFTTT 2087
Db 1799 NVFRETAKQLDPAASVTGSKRQPRTPKGAQPLEDLAGLKELFQTPVCTDKTTHBKTT 1858
QY 2088 TVSITA-----AGTGE--ORQSKLS--PQMOVHODKTLPPAQSSSVGPAKAQOTA----- 2134
Db 1859 KIACRSQDPDVGHPPTTFKPSKSLRKADVEESLALRKRTPSVGKAMTTPKPAAGDEK 1918
QY 2135 -----QPSARPQTPQPSQPAQPEVQTOPEVQTOPTVSS-----HVPSEAQ 2175
Db 1919 DMKAFMGTPVQKLDLPGNLPGSKRWQTPKE-----KAQALEDLAGFKELFOTPGTDK 1971
QY 2176 PTHAQSKPQVAQSQPSQSNVQSGSPVRVQSPSOTRIRPS-----TP 2217
Db 1972 PT-TDEKTTTKIACKS--PQ-----PDPVDTPTASTKQRPKRNLKADVEEFLALRKRT 2022
QY 2218 S-----QLSPGQSOVQTTTSQPIQPIQHTSL--QIP--SOGQPOSQ----- 2256
Db 2023 SAGKAMTTPPAVSDENKINFTVETPVQKLDLLGNLPGSKRQPOTPKEKAQALEDLVCFK 2082
QY 2257 ---QVQSTQTLSSGQTLINVSVSRRPQLOIQOQPOQVIAPV---OLQOOQVILSQI 2309
Db 2083 ELFOTPGHTEESMTDDKITEVCSKSPQESPKTSRSKQRLKIPLVKVDKKEEPLAVSKL 2142
QY 2310 --SQSVVAQIOAQSGVPPQIKL--QLPIQIQSSAVOTHOIQNVVT--VOAASVOEQLOJR 2364
Db 2143 TRTSGETTQHTPTGDSKSIKAFKESPKQILDPAASVTGSRRLRTRKEKARALELDV 2202
QY 2365 VQOL-----RDQOOK-----KKOQOIETK-----EH 2386
Db 2203 FKELFAPGHTTESMTIDKNTKIPCKSPPELDTATSTRKCPTRPRKEVEELSAYER 2262
QY 2387 TLOASNOSEITQOKVWKHNAVIEHLKQ--KKSMTTPAEREENORMIVCNOMKYILDKID 2444
Db 2263 LTQTSQSTHTHKEPASDEG--IKVLQORAKKKPNVPEEESRR----- 2305

```
QY 2445 KEEQAAKKRRRESEVQKRSKONATKLSALLFKHKEQLRA-----ETLKKR 2491
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2306 -----RPRAPKEKAQPLDLAGFTLSGTQESLTAGKATKIPCESPPLEVVDIT 2358
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2492 ALLDKDLQLEVQELKRDILKIKKEDLMOLAQ-----ATA 2526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2359 ASTKRHLRTRVQR-----VOVKEPSAVKFTQTSGETTADKPEAGEDGKIKALKESAKQ 2413
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2527 VAAPCPVPTVLPAPPAPPPSP-----PPGVQHTGLLSTPTLPVASQKRKRREKEK 2578
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2414 TPAPAASTGSRRRPRAPRESAQAIEDLAGFKDPAAGHTTEESMTDDKTKIPCKSSPELE 2473
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2579 DSSSKKKKMISTTSEKTKDKLYICKTPYDESKFYIGDCRQCNWYHGRVCVGLQSE 2638
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2474 DTATSSKRRPRTRAQAVEKKE--LLAVGK-----LTQTS 2506
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2639 AELIDEVVCQOSTEDAMTVLPTLEKDYEGLRVLSLOAHKMAWPFLEPVDPN-DAP 2697
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2507 GE-----TTHYDKPEVGGEGTKA-----FKQPAKRNVDAE 2537
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2698 DYYGVIKERP-----DLATMEERVQR-RYIEKLTFFVAD 2730
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2538 DVIGSRQPRAPKEKAQPLDLASFQELSTQTPGHTTEELANGRAD 2581
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
T21435
hypothetical protein F26H11.3c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T21435
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21435
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-452 <N1>
A:Cross-references: EMBL:Z81515; PIDN:CAB54234.1; GSPDB:GN00020; CBSP:F26H11.3c
A:Experimental source: clone F26H11
C:Genetics:
A:Gene: CBSP:F26H11.3c
A:Map position: 2
A:Introns: 58/1; 89/3; 230/2; 290/2; 334/1
C:Superfamily: bromodomain homology
F;313-368/Domain: bromodomain homology <BRO>

Query Match 2.8%; Score 404.5; DB 2; Length 452;
Best Local Similarity 24.6%; Pred. No. 2.6e-08;
Matches 105; Conservative 82; Mismatches 159; Indels 81; Gaps 10;

QY 2360 EQLQR-VQOLRDQOQKKQOQIEI-----KREHTLQASNQSEIIQKVVMKHNAVI--- 2409
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 ERLKKEIKRRTRMENEACQCGLLTPWKARARPHRAAKPAEVRKEVINPADITLGGD 80
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2410 --EHLKOKKSMTPAEREENQIVMCVNMKYIILDKIDKEEKQAQKRKRRESEVQKRSKQ 2467
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 81 TYDVVKEQKPTESIAATNVSRRRRTSANLSKSEDDR-DKPEQSQTAPKSKERRTSEPPASH 139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2468 NATKLSALLFKHKEQLRAEILKKRALLDKDLQIEVQEEELKRLDIKKKDLMLQAQATV 2527
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 VAFHTPGSATPHDINISIEHCTCQKIFDAS-KLYIQCEL---CARYWHGDCVGVAEQTIL 195
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2528 AAPCPVPTVLPAPPAPPPPPPGVQHTGLLSTPTLPVASQKRKRREKEKSSSKKK 2587
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 196 -----GLEHW-----SCECTEEQER----- 211
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2588 KMISTTSKETKDKLYICKTPYDESKFYIGDCRQCNWYHGRVCVGLQSEAEILIDEYVC 2647
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 -----VKQPALYCVQCPYDDTKFYVGCDCSQGFHPCVGTTRAEAEQAADYNC 262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2648 PQC-----QSTEDAMTVLPTLEKDYEGLRVLSLOAHKMAWPFLEPVDPNDA 2696
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 263 PACTRAEAGYESPASDVSGSSRVSVOLTADYTHVFELLELEHHRMSTFERNVDUNEF 322
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2697 PDYGVIKPEPMOLATMEERVQRRYIEKLTFEVADMTKIFONCRVYNPSDSFPFQCAEVL 2756
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 PDYEKFIKPMDLSTTKKVERTEYLYLSQFVNDVNMOMFENAKTYNPKGNVFKCAETMQ 382
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2757 SFFVQKL 2763
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 EVFDKKL 389
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
A48666
cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: A48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: A48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SCH>
A:Cross-references: EMBL:X65550; NID:g415818; PIDN:CAA46519.1; PID:g415819
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F;29-91/Domain: kinase interaction domain homology <KIH>

Query Match 2.8%; Score 404.5; DB 2; Length 3256;
Best Local Similarity 18.6%; Pred. No. 3.1e-07;
Matches 546; Conservative 378; Mismatches 1049; Indels 955; Gaps 136;

QY 339 IIEEDTENENEKKIWIYSTKVQLAEILDLCK----DYWEALCKLIEEMREIEHRHMDI 394
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 432 VLPTEIHNPEFLTLWLATQVERKIQKDSLSKEKLGTTAGQCMCSGLPGLSS-----VDI 486
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 395 TE-DLTNKARG-----SNKSELAANEEILE-----SIRAKKGDIDN-----VKSPEE 436
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 487 NNFQDSINSEGIPLKRRRVSVFGHLRPELFDENLPNTPLKRGAPTKRKSVMHTPPV 546
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 437 TEKDKNETENDSKDAEKNRE---EFEDQSL-----EKDSDDKTP-DDDPEQKSEVGFDS 488
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 547 LKKIIEQPOPSCKQESGEIHYEVKAQSLVSPAPSPKPTPVASDQRRSCKTPASS 606
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 489 EKNGELSESPGKAGSGASTRIITRLRNPD-----KLSQLKSOQVAAAHAANKLFK 542
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 607 SKSQTEVPK-----RGGERVATCLQKRVISRSQHDILQICSKRRSGASEAN---- 654
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 543 EGKEVLVNSQGEISRLSTKK---EVIIMG---NINNYFKLGOEGKRV--YHNOYSTN- 593
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 655 ----LIVAKSWADVVKLGAKQTQTVIKHGQPSRMNKRQRRPATPKPVGEVHSQFSTGH 710
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 594 -----SFAKNKHQHRDHDKRR--HLAHFKCLTPAGEFK-----WNGSVHGSVKLT 637
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 711 ANSPCTIILCAKTEKHVHPARYVRLNNEFISNQKMDFKEDLSIAEMFKTPYKEQPOLT 770
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 638 ISFLRTITQLENNISPSF-----LHPNASHRANWIKAVQCMCK--PRFALALAI 687
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 771 -STCHTAISNENLLGKQFGTDSGEPLPTSEFSGNFFVFAQAQAQPSD----- 822
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 688 LECAPKPVMLPIWREFLIGHT-----RLHRWTSIERE----- 719
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 823 -KCSASP----PLRRQCIRENGNVAKTPRNTYKMTSLETSTSTETETPSKTVTVNRSR 877
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 720 --EKEKVKK--KEKKOEEEEETM-----QOATWVKYTFPVKHQWVKQKGEYRVTYG 766
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 878 STEFRIQKLPVSKSEETNTEIVECILKRGQATLLQO-----REGENKEI--- 925
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 767 GGWSWISKTHYRFPKLPNGTNNVYRKSLSEGTNNNDENMDESCK-----RKCS---RSP 819
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Db 926 -----ERPEET-----YKENIELKEN-----DEKMKAMKRSRTGWQCAPMSDL 964
Qy 820 KKIKIEDPSE--KDEVGK-----SDAAKADQONEMDISKITEKKDQDVKELLDSDSKP 871
Db 965 TDLKSLPDTELMDKTARGONLLOQDHAHAKAPKSEKGIKTKM-----P 1006
Qy 872 CKE--EPMEVEDDDMKTESHVNCQESSOVVVNVSE-----GFHLRTS----- 911
Db 1007 COSLQP-----EPINWPTHHTQQLKASLGKGVGVKEELLAVGKFTRTSGEHTHREHREAGDG 1062
Qy 912 -----YKKTKT-----SSKLDGLERRIKOFTLEEKORLEKIKLEGIGIKGISTNSKK 961
Db 1063 KSIRTFKESPKIILDPAARVTGM--KKWPRTPKEEAQSLDL--AGFKELFOTPGPSEE 1117
Qy 962 NLSESPVITRAKCEQSDSRMQEOSPNANDOPE-DLIQCSQSDSVLRMSDPSHTTNK 1020
Db 1118 SMTDEKT---TKIACKSPPPESVDPTSTKOMPKRSLRKADVEEFALRLKLTPSAGKAM 1174
Qy 1021 LYPDRVLDVSI-----RSPETKCPKONSTEN-----DIE 1051
Db 1175 LTPPAGDEKDIKAFMGTVPQKLDLAGTLPKSKRLOTPKEKAQALDLAGFKELFQTP 1234
Qy 1052 EKVSDLASRGO-----EPTKSKTKGNDFIDDDKSLASADDIGTLI--CKNK 1095
Db 1235 GHTEELVAAGTKTKIPCDSPQSDPDVDTPTSTKOR-----PKRSIRKADVEGELLACRNL 1288
Qy 1096 KPLIOEESDITIVSSSKSALHSSVPKSTNRD---ATPLSRAMDFEGKLGCDSESNTL 1150
Db 1289 MP-----SAGKAMHTPKPSVGEKDDIIIFVGTVPQKL-----DLT 1323
Qy 1151 ENSSDTVSIQDSEEDMIVQNSNESISEQFTREDDQVEVLEPLKCELVSGESTG--NCEDR 1209
Db 1324 ENLGSKRRPQTPKEEAQALDITGFKELFOTPGHTEEA-----VAACKTTKMPCESS 1376
Qy 1210 LPVKGTEANGKKPSQOKLEBRPNVNCSDQIKLANTDKKNENRESEKKGQRTSTQIN 1269
Db 1377 PPESADTPTSTRROPKTPLEKRDYQKELSALK-----KLQTSGETHTHTDKVP 1424
Qy 1270 GKDNKPKIYAGECLKEISESRVSGNV--EPKVNINKIIPENDIKSLTVKESAIRPFIN 1328
Db 1425 GGEKDSINAFRETAQKLDPAASVTGSKRHPKTK--EKAQPLEDLAGW--KELEFQTP--- 1477
Qy 1329 GDVIMEDFERNSETSKHLLSSDAEGNYRDSLETLPSTKESDSTQTTTSPASCPESNS 1388
Db 1478 ---VCTDKPTTHXKTKIACRSQPD-----PVDTPSSKQSKR 1513
Qy 1389 VNQVEDMEIETSEVKKVTSS-----PITSEESNLNDFIDENCLPINK---NENV 1436
Db 1514 SLRKVDVEEFPALRKRTPSAGKAMHTPKPAVSKEKNIYA--FM---GTPVOKLDLTENL 1568
Qy 1437 NGESKRKTIVITEVTTMTSTVA-----TESKTVIKVERGDKQTVVSSSTENCAKS 1484
Db 1569 TGSKRRLQTPKEKAQALDLAGFKELFOTGRHTEESMT---NDKTAKVACKSSQPDLDKN 1625
Qy 1485 TVTTTTTTVTKLSTPSTCGSVDDIISVKQDSKTVVTTTVDLSLTGTLVTSMTVSKEYS 1544
Db 1626 PASSKRLKTSIG--KVGKVEELLAVGLTQTSGETHTHTHTEPTGDKGSKMAFESPQI 1683
Qy 1545 TRDKVKLMKFRPKTKTRSGTA-----LPSYRK--FVTKTKKSIIFVLPND 1588
Db 1684 LDSAASLTGSKRQLRTPKGSKEVPEDLAGFTLELFTQPSHTKESMTNEKTKTVSTRASQPD 1743
Qy 1589 L-----KKIARGGIREVPYFNYNAPALDINWYPSRPRTFGITWRYRLQTVKS 1637
Db 1744 LVDTPSTSKPQKPSRLRADTEE-----EFLAERKQTPSAG---KAMHTPKP 1787
Qy 1638 LAGVSLMLR--LWASLRWDDMAKVPVPGGGSTRTETSETITTEILIKRRDVGYGIR-- 1694
Db 1788 AVGEEKDINTFLGTVPQKLDOPGNLP--GSNRLQTRKEKAQALELT-----GFREL 1838
Qy 1695 -----FEYCIRKILIC--PIGVP--ETPKETPTPQKGLRSS-----ALRPKRP 1733

1839 FOTPCNDPNTADEKTTKKILCKSPQSDPADTPTNTKQRPKRSLLKADVEEELAFKRLTP 1898
1734 -----ETPKQTPVITETWAAEELEWEIRAF-----AERVE-----KE 1768
1899 SAGKAMHTKAA-----VGEK-----DINTFVGTPEVKLDLLGNLPGSKRRPQTPKE 1946
1769 KAOAVEQQAARLEOQKQPVVIATSTTSPSTSTISPAQKVMVAPISGSVTTGKMWLT 1828
1947 KAKALEDLAGFKELFOTPGHTEESMTDDKITEVSCSKPQDPVKTP-----TSSKORLK 2000
1829 TKVGSFATVTFQONKNFHQTFATVWKQSGNSGVQVQOKVL--GIIPSTGTSTQOFTTS 1886
2001 ISLG-----VGKKEVLPVGLKTOTSGKTTOT--- 2028
1887 FQPTATVITIRPNTSGSGT-----TSNSQVI-----TGPQIRPGMTVIRTP-----L 1929
2029 -----HRETAGDKSIFAKESAKQMLDPANVGTGMERWP-----RTPKEAQSL 2073
1930 QOSTLGAIRTPVMVQPGAPQOVMTQIIRGQVPSTAVSAPNTVSTPGOKSLTSATSIS 1989
2074 EDLAGFKELFOTPDHTEESTTDDKTKI-----ACKSPPPESMDTPTST- 2117
1990 NIOSSASOPPRPOGGOVKLT--MAOLTOLTQGHGNOGLTVVTOGOGOTTGOLQLIP--- 2044
2118 -----RRRPKTPPLGKRDIVEELSALKQLTO-----TTHTDKVPGE 2153
2045 -QGVTLPLPGGOQLM-----QAAMPNGTVQ-----RELFTPLATTATTAS 2083
2154 DGINVFRETAKQKLDPAASVTGSKRQPRTPKGAQPLEDLAGLAKELFOTPVCTDPTTH 2213
2084 TTTTIVSTTA-----ACTGE--QROSKLS--POMQVHQDKTLPQAQSSVGPAPAKAQTA- 2134
2214 EKTTKIACRSPDPVGTPTIFPKQSKRSURKADVEEESLALRRKTPSVGKAMDTKPKAG 2273
2135 -----QPSARPPQOTQPOQPAQPEVQTPQVQTTVSS-----HVP 2171
2274 GDEKDMKAFMGTPVQKLDLDPGNLPGSKRWQTPKE-----KAQALEDLAGFKELFQTP 2326
2172 SEAQPTHAQSKPKPQAAQSQPQSVQSVQSPVQSPQTRIRPS----- 2215
2327 GTDKPT--TDEKTKIACKS--PQ-----PDPVDTASTKQRPKRLKADVEEELALR 2377
2216 --FPS-----QLSPGQSOVQITTSQPIPIQHTSL--QIP--SOGOPOSO----- 2256
2378 KRTPSAGKAMDTKPAVSDKKNINTFVETPVQKLDLLGNLPGSKRQPTQTEKAEALDL 2437
2257 -----QVQSSQTLSGGTLLNQSVSSPSRQQLQIQPQPOVIAVP-----OLQOQOVQ 2305
2438 VGFKELFOTPGHTEESMTDDKITEVSCSKSPQESFKTSRSKQKRLKPLVKVDMKEPLA 2497
2306 LSQI--OSQVVAQIAQOOSGVPOQIKL--OLPQIQOSSAVOTHQIONVVT--VQAASVQE 2360
2498 VSKLTRTSGETTQHTTEPTGDSKSIKAFKESKQILDPAASVTGSRRLRTRKEKARALE 2557
2361 OLQVQOOL-----RDQOOK-----KQOQOIEIKR--- 2384
2558 DLVDFKELFASPGHTEESMTIDKNTKIPCKSPPELTDATSTKRCPTKPRKEVKEELS 2617
2385 --EHTLQASNQSEIIQOVVKNHNAVIEHLQ--KKSMTPAEREENORMIVCQVMQYIL 2440
2618 AVERLTQTSQSTHTHKEPASGDEG--IKVLQRAKKKPNVEEPPSR----- 2664
2441 DKIDKEKQAARKKRESEVQKRSQONATKLSALLFKHKEQLRA-----EI 2487
2665 -----RPRAPKEAQLDLAGFTELSTSGHTQESLTAGKATKIPCESPPLEV 2713
2488 LKKRALLDKLOIEVQBELKRDLIKKEKDLMLQALQ----- 2523
2714 VDTTASTKRLHRLTRVQK-----VQVKEPSAVKFTQTSGETTDADKEPAGEDKIGIKALKE 2768
2524 -ATAVAPCPVTPVLPAPAPPSPPP-----PPPGVQHTGLLSTPTPLPVASQKRKR 2574
2769 SAKQTPAPAAASVTGSRRRPRAPRESAQAIEDLAGFKDPAAGHTEESMTDDKTTKIPCKSS 2828

Qy	1241	KLNTYDKKNNRESEKKGQRTSTFOJINGKONKPKIYLKGECLKEISESRVVVSGNVEPK	1300
Db	297	KKKTTTSKTKTKTKTTPVPTPSSSTTE-SSAPVPTPSSSTTESSAPVTSSTTE--	350
Qy	1301	VNNINKLIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSSTKSHLLSSSDAEGNYRD	1360
Db	351	--SSAPV---TPSSSTTESSAPVTS-----STTESSAPVTSSTTESSAP	394
Qy	1361	SLETLPTSKBDS--TQTTPASCPESNVQVEDMEIETSEVKKVTSSPTISE----	1413
Db	395	VPIPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPV-TSSTTESSAPVTSSTTES	453
Qy	1414	-----BESN--LSNDFIDENGLPIKNNENNGESKKKIVITVITWTST-VA	1457
Db	454	SSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPV	513
Qy	1458	TESKTVLVEKGDKQTVWSSTENCAKSTVTTTTTIVTKLSTPGTGGSDVIISVKEQSKV	1517
Db	514	TPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVPIPS-----SITESSTP	568
Qy	1518	VTVTVDSLTTTGGTLVTSMTVKE--YSTRDKVKLMKFRPKKTRGTGTPALPYRKPVTK	1575
Db	569	VTSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTS	628
Qy	1576	STKKSIVLPNDLKLARKGGIREVPYFNNAKPALDIWYPSPRPTFGITWKYRLQTV	1635
Db	629	STTESSA-----PVPTPSS-----TTE	647
Qy	1636	KSLAGVSLMLRLWLASRLWDDMAKVPVGGGTRTETSEITIT-----TEIKR	1685
Db	648	SSAPVPTP-----SSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVTSSTTE	701
Qy	1686	RDVGPYIREYCIRKILCIQVP-----ETPKETPTQORKGLRSAL-----	1728
Db	702	SSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESS	761
Qy	1729	----RPRRPPTKQTGVIIETVVAEELELWEIRAFERVEKEKAQVQQAKKLEQO	1784
Db	762	APVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSNISS	821
Qy	1785	KP--TVIATSTTS-----PTSTTTSTIPSAQKVMVAPISGVSVTCTKMVLTKVGS--	1833
Db	822	APSTPTPSSSTTESSVPTPSSSTTESS-----APVSSSTTESSAPVPTPSSSN	874
Qy	1834	----PATVTFQONKNPHQTFAWKVGQNSGVVQOQKVLG-----IIPS-----	1875
Db	875	ITSSAPSIPTSSSTESFST-GTVTPSSSKYPCGSQTETSVSSSTETIIVPTKTTTSVT	933
Qy	1876	-----SICTSQ--QTFISFQRTATVIIRPNTSGGGTTSNSQVITGPQIRPGM	1922
Db	934	PSTTTITTVTCSTGNSAGETGSCSKPTVTTV-PTTTTTSVTSSTTIT-----T	985
Qy	1923	TVIRTPLQ---QSTLG---KALIRT-PVMVQAPQOQVMTQIRGQVPSAVS-----AP	1970
Db	986	TVCSTGNSAGETTSGCSPKTITTVPCSTSPSESTTSPTTPTVTVSVTVVTE	1045
Qy	1971	NTVSSSTPGQKSLTSGATSTNIQSS--ASQPPRPOQGV-KLTMQLQLQFGHGGNOGLT	2027
Db	1046	YSTSTKPGGE-ITTFVTKPIPTLYLTIAPTPSVTTVTFPTTITTVTCSTGNSA--	1102
Qy	2028	VVIOGQGTGQLQLIPQGVTVL---PGGQQLMQAAMPNGTVORELFTPLATTATAS	2083
Db	1103	-----GETTSGCS--PKTVTTTVPCSTGGEYITEA-----TTLVTTA----	1138
Qy	2084	TTTTVTSTAAGTGEORQSKLSPOMQVHDKTLPPAQSSSVYGPAKAPOPTAQPSARPOQ	2143
Db	1139	-VTTTVTVTSTGNSAGKTTTG---YTKVSPVTVTVTTLAPS-----	1178
Qy	2144	TOPOSQAPOVQTOPEVQTOVTTVSSHVSEAPHTHAOSSKPKQVAAQSOPOSNVQOSVPR	2203
Db	1179	----APVTPATNAVPTTITTECSAANAAGETTSVCSTKTVSSAGENATPSAT----	1231
Qy	2204	VQSPSQTRIRPST--PSQLSPGQSQVQTT---TSQPIPIQPHHTSLAIPSGQSQSQPOV	2258

Db	1322	--PTVTTAI--PTTVITTTSSVGNNSAGETTGYTTKSIPTYYITLL--IPGSGNKKNYETV	12307
Qy	2259	QSSQTLLSSGQTLNQVSVSSPSPQLQIQPQPQVIAVPOLOQVQVQLSOIQGVVAAIQ	2318
Db	1288	ATAT-----NPSIKTTS-----QLATTASASSVAPV	1315
Qy	2319	AQSGGVPPQIKLQPIQIQSSAVOTHQIONV--VTQAAAS	2357
Db	1316	TSPS-----LTGFLQASGSGVAATYVPSISSTYQAA	1348

```

RESULT 15
T21433
hypothetical protein F26H11.3a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T21433
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21433
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-405 <WIL>
A:Cross-references: MIM:T281515; PTDN:CA804198.1; GSPDB:GNO0020; CESP:F26H11.3a
A:Experimental source: clone F26H11
C:Genetics:
A:Gene: CESP:F26H11.3a
A:Map position: 2
A:Introns: 11/1; 42/3; 183/2; 243/2; 287/1
C:Superfamily: bromodomain homology
P:266-321/Domain: bromodomain homology <BRO>

```

Search completed: September 24, 2003, 01:18:40
Job time : 59.3354 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2003, 19:30:01 ; Search time 19.068 Seconds
(without alignments)
6858.664 Million cell updates/sec

Title: US-09-698-295-10

Perfect score: 14333

Sequence: 1 MYSEEEEDGDAETQDSE.....KLKGFKASRSHNNKLOSTAS 2781

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Description
1	3363	23.5	12830 homo sapien
2	441	3.1	20659 drosophila
3	440.5	3.1	99uif9 homo sapien
4	437	3.0	Q9pu36 gallus gall
5	421.5	2.9	Q9y6v0 homo sapien
6	419.5	2.9	Q9lye5 mus musculus
7	404.5	2.8	P46013 homo sapien
8	401.5	2.8	P08640 saccharomyc
9	396.5	2.8	Q9dgi3 gallus gall
10	391	2.7	Q02817 homo sapien
11	390.5	2.7	Q9qyx7 mus musculus
12	388	2.7	Q9jks6 rattus norv
13	385	2.7	Q9uif8 homo sapien
14	382	2.7	Q60974 mus musculus
15	380.5	2.7	Q99996 h a-kinase
16	379.5	2.6	P39769 drosophila
17	375.5	2.6	Q01484 homo sapien
18	374.5	2.6	P08799 dictyosteli
19	374	2.6	Q12955 homo sapien
20	367.5	2.6	Q14789 homo sapien
21	364	2.5	Q01761 caenorhabdi
22	360	2.5	Q02455 saccharomyc
23	360	2.5	P46100 homo sapien
24	358.5	2.5	Q24742 drosophila
25	356.5	2.5	Q75376 homo sapien
26	355	2.5	P51610 homo sapien
27	354.5	2.5	P12270 homo sapien
28	353.5	2.5	Q9hc84 homo sapien
29	348.5	2.4	Q04893 saccharomyc
30	347.5	2.4	P15205 rattus norv
31	343	2.4	P14873 mus musculus
32	342.5	2.4	Q94833 homo sapien
33	337.5	2.4	P25386 saccharomyc

34	337.5	2.4	2090	1	HFCL_MESAU	P51611 mesocricetu
35	337	2.4	2843	1	APC_HUMAN	P25054 homo sapien
36	333	2.3	2468	1	MAPB_HUMAN	P46821 homo sapien
37	333	2.3	3210	1	CENF_HUMAN	P49454 homo sapien
38	331.5	2.3	2230	1	GOGA_HUMAN	Q13439 homo sapien
39	331.5	2.3	7389	1	BPAL_MOUSE	Q914u6 mus musculus
40	330.5	2.3	1531	1	NFT5_HUMAN	O94916 homo sapien
41	329.5	2.3	3866	1	HRX_MOUSE	P55200 mus musculus
42	328	2.3	2004	1	MOZ_HUMAN	Q92794 homo sapien
43	326.5	2.3	2845	1	APC_MOUSE	O61315 mus musculus
44	326.5	2.3	3562	1	PGCV_CHICK	O90953 gallus gall
45	324	2.3	2476	1	ATRX_MOUSE	O61687 mus musculus

ALIGNMENTS

RESULT 1
ID: FALZ_HUMAN STANDARD; PRT: 810 AA.
AC Q12830;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fetal alzheimer antigen (Fetal Alz-50-reactive clone 1).
GN FALZ OR FAC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95347245; PubMed=7621746;
RA Bowser R., Giambrone A., Davies P.;
RT 'FAC1, a novel gene identified with the monoclonal antibody Alz50,
is developmentally regulated in human brain.';
RL Dev. Neurosci. 17:20-37(1995).
CC -!- SUBCELLULAR LOCATION: DETECTED IN BOTH THE CYTOPLASM AND NUCLEUS
OF CELLS IN THE DEVELOPING CORTEX. IN THE ADULT BRAIN, IT WAS SEEN
ALMOST EXCLUSIVELY IN THE NUCLEI OF NEURONS OF THE NEOCORTEX. IN
THE BRAINS OF ALZHEIMER DISEASE PATIENTS, THE PROTEIN IS LOCALIZED
IN A SUBSET OF AMYLOID-CONTAINING PLAQUES.
CC -!- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN THE FETAL BRAIN.
EXPRESSION IS MUCH LOWER IN ADULT BRAIN AND IS HIGHER IN
NEURODEGENERATIVE DISEASES.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; G01252; G01252.
PIR; G01252; G01252.
TRANSFAC; T04682; -.
GENE; HGNC:3581; FALZ.
MIM; 601819; -.
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005634; C:nucleus; TAS.
GO; GO:0007399; P:neurogenesis; TAS.
InterPro; IPR001965; DDT-dom.
InterPro; IPR004022; DDT-dom.
Pfam; PF02791; DDT; 1.
Pfam; PF00628; PHD; 1.
SMART; SM00571; DDT; 1.
SMART; SM00249; PHD; 1.
PROSITE; PS01359; ZF_PHD_1; 1.
PROSITE; PS50016; ZF_PHD_2; 1.
Nuclear protein; zinc-finger.

RESULT 2
TRX_DROME
ID TRX_
AC P206
DT 01-11

Db 1771 SAEVFEID-----GSLQNHSAVARGMIKCTVCGNRGAT-----VCCNVRSCG 1815
 Qy 1147 ---NSTLENSDTSVSIQDSS-----EDMTVQNSNESISQFRTREODVEVLPLKC 1195
 Db 1816 EHYHYPARSIDCAFLTDKSMYCAHAKGNALKANGSPSV-----TYESNFVSRPVV 1870
 Qy 1196 ELVSGESTGNCEDRLPVKGTENGKPKSQK-----KLEER-----PVN 1234
 Db 1871 EL-----DRKKRLIE-----PARVQFHIGSLVLRQLGAIVPRFSDSYEA VVPIN 1915
 Qy 1235 -KCS-----DOIKLKNTDKNNENRESE-----KKGQRTSTFQINGRDNKPKIYLKE 1282
 Db 1916 FLCRLWYSKPKWKEIVYRTIIONSSTLFDVGRNYTVHDHNPNSKEVOLGMAQI 1975
 Qy 1283 CLAKESBSR---VYSGNVE-----PKVNNINKIIPENDIKSLVAKESAIRPFINGDVIME 1334
 Db 1976 ARWHTSLARSEFLNGGTDSGEFPNPSNC--VPPDQNTTEEPQQQADLLPPELKDAIFE 2033
 Qy 1335 DFNERNSETKSHLLSSDAEGN---YRDSLETLPSTKES--DSTQTTTPSASCPSNSVN 1390
 Db 2034 DL-----PHELLDGISMLDIFLYDDKDLFAISEQSKDGTQAMTS-----N 2074
 Qy 1391 QVEDMETETSEVKKVTSSPITSEESNLSDNDFIDENGLPI--NKNENNVNGESKRTVITEV 1449
 Db 2075 QAQNQQOAGGANSVS-----ICDEDTRNSNTSLG--NGWPASNPVEDAMLSAARN---SSQ 2126
 Qy 1450 TTTTSTVA-----TESKTVIKVEKGDKQTVVSSSTENCACAKSTVTTTTTTVTKLSTPSTG 1502
 Db 2127 VQMLKTLAPKLDGNSAMATAIKRRKLSK-----NLAEGVFLTLSSQORNNKEMATV 2178
 Qy 1503 GSDIISVKQSKTVVTTVDSLTGGLVTSMT--VSKEYSTRD-----KVKLMK 1553
 Db 2179 AGV-----SRRQISSETSVEGATTSGSVRSKSFSTWSAAKRYFEKSEGREEAAKRMQ 2232
 Qy 1554 FSREPKRRTGALPSYRKFTSKPKSIFVLNDLKLKRGKIR-----EVFYFNNAK 1609
 Db 2233 MDG-----VDDSITEFRIIISGDGNLSTAFSGQVKCDRCQCYRNVDAF 2276
 Qy 1610 PALDIWYPSRPRTFGITWRYLQTVKSLAGVSLMLLLWASLRWDDMAAKVPPGGSTR 1669
 Db 2277 QR-----HLPSCSPT-----MSSNETESDVS-----GQGMN 2303
 Qy 1670 TETSETITTEIILKRDVGPYGRFYCKIRKICPIGVETPKETTPQR----- 1720
 Db 2304 NATQISAESLNELOQLLANAGGLNY-----LOSATSFPOVORLGLSGOFLG 2350
 Qy 1721 KGLRSSALRPK-----RPETPKGTPVLIETWVAEELELW--EIRAFARV-- 1765
 Db 2351 QGLQLOLQPSLNGEFLSOPNATOANT-----DDLQIYANSLOGLAANLGG 2399
 Qy 1766 -----EKEKAQAVEQQAQRLEQOKRPVIATSTTSPTSTSTTSISPAQKVM 1811
 Db 2400 GFTLAQPTVTPAPAQQLIAVSTNPDGTQQFTQIPQTMQATTTPTATYQTLQATNTDKIM 2459
 Qy 1812 VAPISGSVTTGKVLV--IKVGSPTATVQONKFNHOTATWVK-----QGQSNAGVVO 1864
 Db 2460 L-----PLTAAGKPLKTVAKAQAQAAVKORQLKSHGVKPIQAKLPHPQHQOQOQO 2514
 Qy 1865 VQQ--KVLG-----IIPSSTGT-----SQQTTSF----- 1887
 Db 2515 VQQPTVWGQNLQPOLLFQSSGTQAPILPQAQPNIIISFVYDGSQGPLOYISIP 2574
 Qy 1888 -----QPR--TATVTIRNPNTSGSG-----TTSNS---QVITGQPIRPM 1922
 Db 2575 TAGYKPKQPQPTATPTLTAPGAGATYLTQDASGNLVLTTTPSNSGLQMLTAQSLQAQ 2634
 Qy 1923 TVIRTPLQOSTL-----GKAIIRTPVMVQPAQVVM----- 1956
 Db 2635 QVIGTLIQPOTIQLGGGADGN-----QPGNQOPLILGGTGGSSGLLEFATTSFQV 2685
 Qy 1957 IIRGQPV-----STAVSAPNTVSST-PG-----QKSLTSATSTSIQSSAQPPRPQ 2003
 Db 2686 ILATQPMYGLTETIVQNTVNSSQQFVSTAMPGLMSQNASFSATTTQVFQASKIEP----- 2740

Qy 2004 GOVKLTMAOLTQITQHGNGOGLTV-----VIOQGGQ--TTGQLQLIPQGVTVLPGGQOL 2057
 Db 2741 -IVDLPAQVVLNNTGDASSAGTFLNAASVLOQOQDDTTTQI----- 2782
 Qy 2058 MQAAMPNGTVQRELFPTPLATTATTASTTTTSTTAAGTGEQQRKSLSPOMOVHODKTL 2117
 Db 2783 ----LQANFQ--FQSPTSSGASTSMDYTSPVMVTA-----KIPVTOIKRTNAQA 2828
 Qy 2118 PAQS--SSVGPAKAQPO-----TAQPSARPQPTQPOSPAQEVQTOPEVOTQTTVSSH 2169
 Db 2829 KAAGISGVKVPPOPVQVNVKVLPTSIVTQOSQOVQVKNLKSQVYKGAASGTGTTGCA- 2887
 Qy 2170 VPSEA-----OPHTAQSKEPV-----AASQPSQNVQGSQSPV 2203
 Db 2888 PPSIASKPLQKKTMNIRPIHKLEVPKVMKPTPKVQONQNHSLQOQOQOQOQLOQOIPAV 2947
 Qy 2204 V--QSP-----SOTRIRPS-----PSQLSPGQSOVQVQTTTSOPI---PIQPH 2241
 Db 2948 VNVQVPKVTISOQRIIPAQTOOQLOQAQMIHPQOQOPLQOQOVQVQPSMPIITLAEAPV 3007
 Qy 2242 TSLQIPSQGPQSQPOVQSSQTLSGGQTLNQSVSPPSR----- 2281
 Db 3008 VQSQFVMEPQALQEQELANRVQHFSTSSSSSSNSCSLPTNVVNPQQQAPSTTSSSTRP 3067
 Qy 2282 ----POLQIQPOP-----QVTAQPOLQOQVQ--VLSQIOSVVAQIOAQSGVPOQIK 2329
 Db 3068 TNRVLPQMQRQEPAPLSNECPVVSPTPKPVEQPIIHQMTSASVSKCAQKSTLSPV- 3126
 Qy 2330 LQPLQIOGSSAVQTHQIQNVTVQAASVQEQQLRVQOQLRDQOQKKQOQIEIKREHTLIQ 2389
 Db 3127 --YEALKVSSVLES--IYPDVTMDA--ILEEPVTVQSIYTEGLYEKNSPGESKTQLLL 3180
 Qy 2390 ASQSEIIQKVYMKHNAVIEHLKQKSM--TPAEREN-----ORMIVCN- 2433
 Db 3181 QQQQREQLNQLVNGVLLDKHTFQVEPMDTVYREEDLEEEDEDEDFSLKMTSACND 3240
 Qy 2434 ----QVMKYLIDKIDKEE-----KQA 2450
 Db 3241 HEMSDSEEPVAKDKISKILDNLNDLDDCADSIATATMEVDVDSAGYQOMVEDVLATTAQS 3300
 Qy 2451 AKRKRREESVEQKRKONATKLSALLFKHKEQLRAELTKRALLD-KDLQIEVQEELKRD 2509
 Db 3301 APTEEEGALTAATAVEAAATVINEMADAH-----VLDLKLQQLNGVLELELR 3347
 Qy 2510 LKTKKEKDLMLQAATAFAAPCPVPTVLPAPPAPPPPPPPPG 2553
 Db 3348 ---KEQRTVSQEQSAAIIVP--TAAAPPPQPIQEPKMTG 3386

RESULT 3

BA2A_HUMAN STANDARD; PRT; 1878 AA.
 ID BAZA_HUMAN Q9UIF9; O005336; O15030; Q96H26;
 AC Q9UIF9; O005336; O15030; Q96H26;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Bromodomain adjacent to zinc finger domain 2A (Transcription
 DE termination factor-1 interacting protein 5) (TF-1 interacting protein
 DE 5) (Tip5) (hWALP3).
 DE BAZ2A OR TIP5 OR KIAA0314.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=20130112; PubMed=10662543;
 RA Jones M.H., Hamana N., Nezu J.-I., Shimane M.;
 RT "A novel family of bromodomain genes";
 RL Genomics 63:40-45(2000).
 RN [2]


```

Db 103 SANPGSNLKPDPILLSQFSGG-----QYPLNGILGSRQPS-----PS 140
Qy 1016 HTTN-----KLY-----PKDRVLDVVSIRSPETCKPQNSTENDIEKVSDLASRGQE 1063
Db 141 HNTNLRAGSQKFWANGTHSPMGLNFDQSLEYD---SFPDQN-----FEEVCS-----GTH 187
Qy 1064 PTKSKTKGNDFDDDSKLASADDIGT-LICKNKKPLIOESDRIVS-----SSKSALH 1115
Db 188 PDEAAEKEMTSVVAEN-----GTGLVCSLE---LEEEQPELKWCGNGSVPSVESLH 236
Qy 1116 SSVPKSTNDRDAPLSRAMDFEGLGCDSSNSTLENSDTSVSIQDSSSEEDMIVQNSNES 1175
Db 237 QEVSVLVPD---PTVSCLD-----DPSHLPDQLEDTPLS-----268
Qy 1176 ISEQFRTRQODVEVLEPKELVSGESTG-----NCEDRLPVKTEANGKKPSQOKKL 1228
Db 269 -----EDSLEPFNSLAPEVSGGLYGIDDTLMGAEDKLP-----L 304
Qy 1229 EERPVKCSQDIKLKNTTDKKNENRESEKKGORTSTFOINGKDNKPKIYLKGECLKEIS 1288
Db 305 EDSPVISALDCPSLNAT-----AFSLADDSQ-----332
Qy 1289 ESRVSVGNVEPKVNNINKIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSETKSHL 1348
Db 333 -----TSTSFASPTSPVVLGESVLQD-----NSFD-----L 359
Qy 1349 LSSDARGNVRDSLETILPSTKESDSTQTTTPSASCPSNSVNQVEDMEIETSEVKKVTSS 1408
Db 360 NNGSDAE---QEEME---QSSDFPPLSTOPA---PDQSSTIQLHP---ATSPAVSPPTS 407
Qy 1409 PITSEESNLNDRIDENGLPINKNNVNGESKRKTIVITEVTTMTSTVATESKTVIKVEK 1468
Db 408 PAVSLVSPASPEIS-----PEVCPAASTVV---SPAVSFVS 443
Qy 1469 GDKQTVVSSSTENCASKTIVTTTIVTKLSTPSTGSGVDIISVKEQSTVTTTIVTDSLTT 1528
Db 444 PASSAVLPAVSLVPLTASVTSKRASPVTPSPAAFPASPANKDVSSFLETTADVEEITG 503
Qy 1529 TGGILVTSMTV-SKEYSTRDKVL---MKFSRPKTKSGT-----AL 1566
Db 504 EGLTASGSDVWRRARIAPPEEVRPLPLOGHWRREVRIKKGSHRWGETWYGPCGKRMQKF 563
Qy 1567 PSYRKFTVKSTKSFIFLNPDLKLARKGGIREVPYFNYNAPALDINWYPSPRPTFGI 1626
Db 564 PEVIKLSRLNVHSV-----RRE-HFSF-----SPRMPVGD 593
Qy 1627 TWRRLQTVKSLAGVSLMLRLWLASLRWDDMAKVP---PGGSGTSTETSET-----1675
Db 594 FFEER-DTPEGLQ-----WVLSAEIIPSRIOAITGKRGRPRTERAKTKEVPKVK 643
Qy 1676 -----BITTEIIKRDVGYPYIGIRFYCIRKIIICPIGVPTPKETPTPQKGLRSSAL 1728
Db 644 RGRGRPPKVKITELLNKTDNR-----LKKLEAQETNEEDKAKIAKSKKMRQKVQ 695
Qy 1729 R-----PKRETQKQGPVITETWVAEELELEWEIRAFARVEKEKAQAVEQ 1775
Db 696 RGECLTTIQQOARNRKOETSLAKHKEAKKSKEKEGKTKQEKLEKVKREKKEVKM 755
Qy 1776 QAKKRL-----EQQKPTVATSTTSPTSTST-----1803
Db 756 KEKEVTRAKPACKADKATLQATRLREERQKQOMILEEMKKPTEDMCLTDHQPLPDFSRVP 815
Qy 1804 -----ISPAQKVMVAPI---SGSVTTGTGM-----1825
Db 816 GLTLPAGAFSDCLTIVEFLHSFGKVLGDFDPAKDVPSLGLVLQEGLLCQGDLSLGEVDLIVR 875
Qy 1826 VLTTKVGPATVTFQONKFNHQTFAWVKQGSN-SGVVVOQKVGLIIPSTGTSQ-QT 1883
Db 876 LLKAALHDHPGFPSCQSLKILGEKVSEIPLTRDNVSEILRCFLMAYGVEPALCDRLRQTQP 935
Qy 1884 FTSFQP-RTATVTIRPNTSGSGT-----TSNSQVITGPQIRPGMTVIR 1926
Db 936 FOQPPQOKAAVLAFVVELNGSTLIINEIDKTLSEMSYRKNKWIVBGLRR-----LK 990

```

```

Qy 1927 TPLQOSTILGKAIIRTPVMVQPGAPQOVM-----TOIIRQOPVSTAVSAPNTVSTSPGQKS 1981
Db 991 TVLAKRT-GRS-----EVEMGRPEECLGRRSSRIMEETSGMBEEEEESIAAIPGRR- 1042
Qy 1982 LFSATSTSIQSSASOPRPPQOGQVKLTMAQILTQTQHGNGOGLTVVLOGGQTTGQLQ 2041
Db 1043 ---GRRDGEVDATASSIPELERQIEKLSKRQLFFRKLLHSSQMLRAVSLGDRYRRYW 1099
Qy 2042 LIP---QGVTVLPGCGOOLMQAAMPNGT-----VQRELTPLATTATTASTTTTSTTA 2093
Db 1100 VLPYLAGIFEVEGTEGLVPEEVIKKEITSLKVAHAASLNPALFSMKMELAGSNTTASSPA 1159
Qy 2094 ACTGEORQSKLSPQMGMVHODKTLPPAQSSSVGPAKAQPTQAQPSAPQPTQTPQSPAQPE 2153
Db 1160 RARGRPRTK-----PGSMQPRHLKSPVRGODSEQQAQLOPE 1197
Qy 2154 VQ-----TQPEVOTQTTVSSH---VPSEAOPTHAQSSKPOVA-----AQSOPOSNNVG 2198
Db 1198 AOLHAPAPQPOLQLOLQKHGFLEQEGSPLSLGOSQHDLSQSAFLSWLSQTSQSHLLS 1257
Qy 2199 QSPVRVQSPQTRIRPSTPSQLSPGQSOVQTTTQIP-ITQHTSLQIPSGQSPQSP 2256
Db 1258 SS-VLTPDSSPKLDPA-PSQ--PPEPEPDEAESPDQALWENLSAQMPCNAAPTTPP 1313
Qy 2257 VQVS-----STQTLSSGOTLNOVSVSSPSRPOLI-----OQPO-PQ 2292
Db 1314 AVSEDOPTSPQOLASSKPMNRPSAANPCSP-VQFSSTPLAGLAPKRRRAGDCEMPOSPT 1372
Qy 2293 VITAPVLOQO---VOVLSQISQSVVAVQIAQQSGVPOQ-----IKLQLP 2333
Db 1373 GLQOPKRRRPPSKFKQMEQRYLTQLTAP---VPPMCSGWWIRDPPEMLDAMLKALHP 1430
Qy 2334 IQIQOSSA---VOTHQ---IQNVVTVQAAS---VQEQLQRVQ---LRDQOQKKQOQIEIKR 2384
Db 1431 RGIREKALHKLHNRDFLOEVLCLRPSADPIPEPRQLPAFQEGIMSWSPKETYETDL-- 1488
Qy 2385 EHTLOASNOSEITOKQVVMKHNAVIEHLKOKKSMTPAEEREENORMIVCQVMKYILDKTD 2444
Db 1489 -----AVLQWVEELEQVIMSDLQIRGWTCPSPDSTREDLAYCEHLS-----D 1531
Qy 2445 KEKQAAKKRKEESVEOKRSKONATKLSALLFKHKEQLRAETLKKRALLDKDLQIE-VQ 2503
Db 1532 SQEDITWRGREGRLAPQRT-TNPLDLAVM-----RLAALEQNVERRYL 1576
Qy 2504 EELKRDLIKKEKDLMLQALQATAVAAPCPVTVLPAPPAPPPPPPPGVQHTGLLSTP 2563
Db 1577 EPLWPTHEVVLEKALLS-----TPNGAPEGTTTETISEYETP 1612
Qy 2564 TLPVASQKRKEEKDSSSKSKKKMISTTSKE-----TKKDTKLYCICKTPYDESKFYI 2618
Db 1613 RIRVWROTLE-----CRSAAQVCLIGQLERSTAWKSVNKVTCVCRKGDNDFFLL 1665
Qy 2619 GDCRCQWYHGRVCVGIQSEAEELIDE--YVCPQC--QSTEDAMT-----2658
Db 1666 LCDCCDRGCHYC---HRPKWEAVPEGDWFCVCLAQOQVEGETQKPGFPKRGQKRKSGY 1722
Qy 2659 -----VL-----TPLTEKDY--EGL--KR-----VLRL 2678
Db 1723 SLNFSEGDGRRRRVLLRGRESPAAGPRYSEEGSPSKRRRLSMRNHHSDJTFCEIILMEM 1782
Qy 2679 QAKHMAWPELEPDPNDADPYGYGVKPEMDLATMEERVQRRYYEKLTEFVADMTKIFDNC 2738
Db 1783 ESHDAWPELEPVPNRLVSGYRRIKKNPDMFSTMRERLLRGYTSSEFAADALLVFDNC 1842
Qy 2739 RYVNPSPSPFYQCAEVLESFFVQKLKGFKASRSIN 2773
Db 1843 QTFNEDDSEVGKAGHIMRRFFESWEFYOQKQAN 1877

```

RESULT 4
PCLO_CHICK
ID PCLO_CHICK STANDARD: PRT: 5120 AA.

Q9PU36;
 28-FEB-2003 (Rel. 41, Created)
 28-FEB-2003 (Rel. 41, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Piccolo protein (Aczonin) (Fragment).
 GN PCLO OR ACZ.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99439764; PubMed=10508862;
 RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
 RA Killmann M.W.;
 RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
 RT zones, shares homology regions with rim and bassoon and binds
 RT profilin.";
 RL J. Cell Biol. 147:151-162(1999).
 CC -!- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).
 CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
 CC synaptic junctions (By similarity).
 CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -!- SIMILARITY: Contains 2 C2 domains.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y19187; CAB60725.1; -
 CC HSPB; P04410; 1A25.
 CC GO; GO:0045202; C:synaptic junction; ISS.
 CC GO; GO:0003509; F:calcium ion binding activity; ISS.
 CC GO; GO:0003544; F:calcium-dependent phospholipid binding acti. . . ; ISS.
 CC GO; GO:0003522; F:profilin binding activity; ISS.
 CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 CC GO; GO:0016080; P:synaptic vesicle targeting; ISS.
 CC InterPro; IPR000008; C2.
 CC InterPro; IPR001478; PDZ.
 CC Pfam; PF00168; C2; 2.
 CC Pfam; PF00595; PDZ; 1.
 CC PRINTS; PR00399; SYNAPTOTAGMN.
 CC SMART; SM00239; C2; 2.
 CC SMART; SM00228; PDZ; 1.
 CC PROSITE; PS00499; C2_DOMAIN_1; 1.
 CC PROSITE; PS50004; C2_DOMAIN_2; 2.
 CC PROSITE; PS50106; PDZ; 1.
 KW Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
 KW Repeat.
 FT NON_TER
 FT DOMAIN 258 357 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
 FT P-A-K-P-Q-P-Q-P-X.
 FT 2N_FING 368 392 C4-TYPE (POTENTIAL).
 FT 2N_FING 836 859 C4-TYPE (POTENTIAL).
 FT DOMAIN 2324 2343 POLY-PRO.
 FT DOMAIN 4414 4493 PDZ.
 FT DOMAIN 4627 4726 C2 DOMAIN 1.
 FT DOMAIN 5003 5094 C2 DOMAIN 2.
 KW SEQUENCE 5120 AA; 560751 MW; A658D9891B65B412 CRC64;

Query Match 3.0%; Score 437; DB 1; Length 5120;
 Best Local Similarity 17.6%; Pred. No. 4.9e-08;
 Matches 455; Conservative 385; Mismatches 945; Indels 798; Gaps 99;
 QY 415 EELESIRAKKGDIDNVKSPETEEDKNETENDSKDAENRREFEEDQSLKSDSDTTPDD 474
 DB 965 EKTLSADKIQGG----IQKEDAKSKOGKLFKTPSADKIORVSKOEDSRLOOTKLTTPSS 1020
 QY 475 DPEQKSEYGVDFKSEKNGELSPGAGKSGASTRIITRLRNPDSKLSQKLSQQVAAAA 534
 DB 1021 DXILHGVQKEDIKFQEA--KLAKIPSADK-----ILHRLQKEDFKLOQMK---MAKA 1067
 QY 535 HEANKLFKEGKEVLVNVNSOGESIRLSSTKKEV---IMKGNINNVFKLGQEGKYRYVNIQYS 591
 DB 1068 LSADKIQPEAKQEDVOLQEVRLSKAVSADKIOHQIKDLNLQHVKTSEKTSSEVKIEQAOK 1127
 QY 592 TNSFALNK-----HQ---HREDHDKRRHLAHKFKCLTPAGEFKWNGSVH 631
 DB 1128 ESKLQODKLPKTLSEDKIPATVSSDHKLLSKSEEDKPELLEK--STPHPKDK----- 1179
 QY 632 GSKVLTITSLRLTITOLENNIPSSFLLHNPWASHRANWIK-----AVQMCSPREFALA 684
 DB 1180 KEQIPAEITGHITQEKAVEAPCDKLH---EKKQEDVKKEDLTTCIPQWVSKP----- 1229
 QY 685 LAILECAVPVVMPLIWRFLGHTLRHMTSIEEREKEKVKYKKEKQEBEETMQQATWTK 744
 DB 1230 ----EKAEEKTPVPV-----SRLPRSDHVE-AVREKIEKDDKSDTSSSQOQ----- 1272
 QY 745 YTFPVKHQVWKQGEERYVTGCGWSWISK-----THV---YRFVKLPNGTNNVNRKS- 795
 DB 1273 -----KSPQGLSDTGYSSDGISSSLGSEIPSHIPSDEKDLPREPSQKDTISOESP 1321
 QY 796 -----LEGTKNNNDNN-----DE-----SDRKKSRSPPKIKIEPDSEK 830
 DB 1322 PPSDLAKLESIVLSILEAQASTLTDEKSVKRKELYETVSEQTKDQHKTKPLPVTESYS 1381
 QY 831 DEVKGSAAKADQNDMDISKITEKDDQVVELLSDSDKPKCKEPEME-VDDDMKMTESHV 889
 DB 1382 SDEEDLEATQEGERTIAADSKGASSTQDYKE-EDGNDTPARRQYRSDVESSESNP 1440
 QY 890 NQESSQVDVNVNVEGFHLR-----TSYKKTKKSKLDGLLERRIK 930
 DB 1441 VPRKRRASVGSSSSDEYARDSSQSGDEEDFTRKOLIEMSADEDSGEDDEFIRNOLK 1500
 QY 931 QFTLEEKQRLEKIKLEGGIK-IGTKTSTNSKNLSPEVITKAKQCSDSMRQ-----EQ 985
 DB 1501 EISVTESSQKKEEVKSKA--KGTVGK-----HRRMARKSSAGYDEDAGRHSHWDD 1548
 QY 986 SPNANNQPEDLIQGCQSQSDSVLRMSDPSTHTNKLYPKDRVLDVVISRSPETKCPKQNS 1045
 DB 1549 DDETDESPEPKYRETKSQDGEELAISSGG-----GLRRFKT-----IE 1587
 QY 1046 IENDIEEKVSLASRGQEPKSKTKGNDFIDDSKLASADDIGTLCKNKKKPLIQEESDT 1105
 DB 1588 LNSTITSKYSE-----TPQOKGILYFDEEPPEL-----EMESUTDSPEDR 1627
 QY 1106 IVSSSKSALH-----SSVPKSTNDRDAPPLSRAMDFEGKLGCDSESNSTLSSND 1155
 DB 1628 SRGEGSSSLHASSFTPTGTSFTSVSSLDSDSSPSHK-----KLGESKQQRKARHSH 1681
 QY 1156 ----TVSIQDSSSEDMIVQNSNISISQFRTRE-----STFOI----- 1268
 DB 1682 GPLLPTIEDSSBEEL-REEEELLKQEQRELEQQOQKSSSSKKSKKDKDELRAQRERR 1740
 QY 1185 -----QDVEVLEPLKCELYSGE-STGNCEDRLPVKGTENGKPSQOKKLEERP 1232
 DB 1741 PKTPPSNLSPIEDASPTTELROAAEMELHRSCSEYSPSIESDPEGFISPEKIIIEVQK 1800
 QY 1233 VNKSDQIKLKNITDKKNENRESEKKGORT-----STFOI----- 1268
 DB 1801 VYKLPVAVSLYSPDTEK-LIGALKESGGKTLKSAEEVVEEMHKTHTKSKSFQIASEKDE 1859
 QY 1269 -----NGKDNK-----PKI 1277

Db 1860 VFEKESLYGMLIEDYIESLIEDTYNGTVDITNLAMRQDESNEYIQORGKERRASEQI 1919
||: |
Qy 1278 YLGECLKEISERVVSGNVEPKVNNINKIIPENDIKSLTVKESAIRPFINGVIMEDFN 1337
||: |
Db 1920 YDEPQKTTDLQEDYY---SVPE---LCSIVPOEDI---VSSYIIP-ESHEIVLDST 1967
||: |
Qy 1338 ERNSSEFKSHLSSDAEGNROSL--ETLPSTKESDSTQTTTPSA--SCPESNSVNOVE 1393
||: |
Db 1968 VTSTTEKQLL---DAAAYEELMKQRMQLTPGSSPTQPTSDLAPTSMDKVVSSIGEIA 2023
||: |
Qy 1394 D---MEIETSEYKVTIS-----SPITSEEE 1415
||: |
Db 2024 DSTSLSTSTSAISDVSSLSIALSIPDVKITQHFTAEEIEDEYLDIYAREIOEIIISHET 2083
||: |
Qy 1416 SNLSNDIDENGLPINKNNVNGESKRKTVITEVTWTSTVAV-----ESTVIKVEK 1468
||: |
Db 2084 SMLTYSEVSGEASILPSDTAS-----LTSSTSSVCITDSSSPIDSATTGYVDT 2132
||: |
Qy 1469 GDKQTVSSSTENCAKSTVTTTTTTVTVKLSTP--STGGSVD-----IISVKE 1512
||: |
Db 2133 SDAVSKLVDSEDIIAQVPFTSTEEYSEVSMYESVAGATTKPAIASDMDTVHQAAVCLPE 2192
||: |
Qy 1513 QSKTVVTTVTDLSLTTGGTLVMTVSKYSTRDKVKLMKFSRPPKTRSGTALPSYRKF 1572
||: |
Db 2193 TAPSVFTTWTIKPKQYASDTIYDISTAE---KDAARKMK-----STVETGIIKI 2239
||: |
Qy 1573 VTKSTKKSIFVLPNDLLKLARKGGIREVPYFNVA---KPADIWPSYPPRPTFGITW 1628
||: |
Db 2240 HHDSHKLSJL---DMTRINLNGATSEQPLCLVASVKEPASEIPAVPTPRV---VSK 2292
||: |
Qy 1629 RYRLQTVKSLAGVSLMLRLWASLRDMDMAAKVPPG----- 1664
||: |
Db 2293 TSTVSMPPSSAPALTSKVFSLFRSSLDSPAQSPSPPPPPPPPPPPPLPPPLPKAIPY 2352
||: |
Qy 1665 -----GGSTRITSTETITTEIIRKRDVGPYGRFEYCIRKICPIGVPETPRETPT 1717
||: |
Db 2353 KKSQIQAPMATAPVPLVTSVATLESAAV-----LKNHVVPVTKTVTPTPPPV 2402
||: |
Qy 1718 PORKGLRSSAL---RP---KRPEPTKGPVLIETWAAEELEWEIRAFARVERKEK 1769
||: |
Db 2403 PPKPSSIPAGLVFHSRTEVTKPIAPKPAVPL----- 2436
||: |
Qy 1770 AQAVEQAKKLEQOKPTVIATSTSTSTSTSTISPAQKVMVAPIS----- 1816
||: |
Db 2437 PIAVHKPA-----ETQPKPIGLSLTS--SMTNLVSSAEYKIASPTSPSPHNSKSSPRL 2489
||: |
Qy 1817 --GSVTTCTKMVLTKVGSFA-TVTFOQKNFHOTFATVWKOGOSNGVQVQOKVLGII 1873
||: |
Db 2490 TKPSOETVVITLSEPGTPEAIT-----SOAVTSWLEAPSKQEIQPMQPIF--- 2539
||: |
Qy 1874 PSSGTGTSQOTFTSPQRT-----ATVTRPNTSGSGGTTSSNQVITGPQIRPGMT 1923
||: |
Db 2540 -----TSSMKAVEIQSMADQSMYITGALQITPITQTSTFEKVPSSKSEAVTEVAKTAS 2594
||: |
Qy 1924 VIRTPLOQSTLGKAIIRFP-----VMQVP----- 1947
||: |
Db 2595 VVKRPVPSVGLSVTITIPPEPIVISDQPRYRENGRPHFGDVIDLRLTKVDIEMRDCS 2654
||: |
Qy 1948 -----GAPOQVMTQIIRGOPYSTAVSAPNTVSST-----PGQKSLTSA-----TSTSN 1991
||: |
Db 2655 MDLSAVSMRARQMTSDTSRPSVTQVPAIINLSTACVADPSLSIVTEIVAVMTCATV 2714
||: |
Qy 1992 QSSAQPPRPQGGVKLTMAQLTOL-----TOGHGG 2022
||: |
Db 2715 SYSASTDSLVDLGHAMTTPQLTTSKHFEPAIRVTSQPPFVSRDEVPINLSLGTSAHAV 2774
||: |
Qy 2023 NQGLT--VVIQOGOTTQQLQLIFQGVTVLPGCQQLMQAAMPNGTVQRFLETLATAT 2080
||: |
Db 2775 TWAATKPTVVPVPSVINGWTDL-----STSQEPMEIGAVD--LSTTKSHRTV 2820
||: |
Qy 2081 TASTTTTTVSTTAAGTGBQRQSKLSPOMQVHQDKT--LPPAQSSSVGPAKQAPQATPFA 2138
||: |

Db 2821 TWDETTSGIITTVIEDDEKPVDLTAGRRVCCDMVKLPFGRSC-----TAQ--- 2867
||: |
Qy 2139 RPQPTQPOSPAQPEVQTOPEVQTOPTVSSSHVSEAPTHAQ-----SSRPQVAAQSPOS 2194
||: |
Db 2868 --QPPT-----TLPEDRFYR-DDHYOYDRSGSYGYRGMGMKPSMSDNLSEA 2913
||: |
Qy 2195 NV---QGSPVVRVQSPQTRIRPSPQSLSPGQSQVQVQTTTQPIP-----IQPHTSLOIP 2247
||: |
Db 2914 GLFAVKSNSFDYQVGATDAADVLTSGRVTSGEVMDSYSSKTTGPPETRQVVISGIGISTP 2973
||: |
Qy 2248 SOGQPO---SQPVQSSQSTQTLSSGOTLNOVSVSPSRPQIQIQ-----POP 2291
||: |
Db 2974 QTSQARMWSSLSPPFAGSVLRSSNGVYSSVATPIPTFAITTPGSGIFSTVTRDPLPL 3033
||: |
Qy 2292 QVT-AVPLQ---QQVQLVSIQSVVQVQIAQOQSGVQPOIKLQPLTQIQIOSSAVQTHQIQ 2348
||: |
Db 3034 QTDISVPSLSTLQONQPLPRSYSLTMAAEKDASTLTDIETGLPLTLESIAITED--T 3090
||: |
Qy 2349 NVTVQAAASVQEQLRVQQLRDQ-----OQKKQOOIEIKREHTLQASNQSEIIQKV 2402
||: |
Db 3091 NLIPATASEVYTDV-----IEDEVALIIAPEEGKQOQDLER-----ELLE--- 3131
||: |
Qy 2403 MKHNNAVIEHLKOKKSMTPAEREENQRMIVCNVMKYILDKIDKEEKAAKRRKRESVQ 2462
||: |
Db 3132 -----LEKIKQOR-----FAEELEWERQEIQREREQEKFMVQ 3163
||: |
Qy 2463 KRSKONATKLSALLFKHKFOLRAE--ILKRALLDKDLQTE-----VOEELKRDLIKKEKD 2517
||: |
Db 3164 KKEELQSMKHHLLFOQEEERQAYMMRQETLAAQOQLQLEQFOQLQOQLHQLEQKIQ 3223
||: |
Qy 2518 LMQ 2520
||: |
Db 3224 IYQ 3226

RESULT 5
PCLO_HUMAN
ID PCLO_HUMAN STANDARD; PRT: 5147 AA.
AC Q9Y6V0; Q60305; Q9BVC8; Q9UIV2; Q9Y6U9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Piccolo protein (Aczonin) (Fragments).
GN PCLO OR ACZ OR KIA00559.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-759 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99439764; PubMed=10508862;
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA Kilimann M.W.;
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
zones, shares homology regions with rim and bassoon and binds
profilin";
RL J. Cell Biol. 147:151-162(1999).
RN [2]
RP SEQUENCE OF 552-4404 FROM N.A.
RC Kraemer J., Wollam C., Wohldmann P., McGrane B.;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro";
RL DNA Res. 5:31-39(1998).
RN [4]

RP SEQUENCE OF 4405-4439 FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Heiton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5]
 RP SEQUENCE OF 4405-5147 FROM N.A.
 RA Kalicki J., Elliott G.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -!- SUBUNIT: Interacts with Rabac1/pral and profilin (By similarity).
 CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
 CC synaptic junctions (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing: Named isoforms=2;
 CC Comment-Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q9Y6V0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
 CC VSP_003926, VSP_003927;
 CC Note-No experimental confirmation available;
 CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -!- SIMILARITY: Contains 2 C2 domains.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Y19188; CAB60727.1; -;
 DR EMBL; AC004903; AAD20936.1; -;
 DR EMBL; AC004886; AAD21789.1; -;
 DR EMBL; AB011131; BAA25485.1; -;
 DR EMBL; BC001304; AAH01304.1; -;
 DR EMBL; AC004082; AAB97937.1; -;
 DR FIR; T00634; T00634.
 DR HSSP; P04410; 1A25.
 DR Genew; HGNC:13406; PCLO.
 DR MIM; 604918; -;
 DR GO; GO:0005856; C:cytoskeleton; NAS.
 DR GO; GO:0045202; C:synaptic junction; ISS.
 DR GO; GO:0005509; F:calcium ion binding activity; ISS.
 DR GO; GO:0005544; F:calcium-dependent phospholipid binding acti. . .; ISS.
 DR GO; GO:0005522; F:profilin binding activity; ISS.
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO; GO:0016080; P:synaptic vesicle targeting; ISS.

DR InterPro: IPR000008; C2.
 DR InterPro: IPR001565; Synaptotagmin.
 DR PRINTS; PR00360; C2DOMAIN.
 DR PRINTS; PR00399; SYNAPTOTAGMN.
 DR SMART; SM00239; C2; 2.
 DR PROSITE; PS00499; C2_DOMAIN_1; 1.
 DR PROSITE; PS00004; C2_DOMAIN_2; 2.
 KW Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
 KW Repeat; Alternative splicing.
 FT NON_TER 1 1
 FT DOMAIN 400 465 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
 FT 2N_FING 499 523 P-A-K-P-Q-P-Q-P-X.
 FT 2N_FING 969 992 C4-TYPE (POTENTIAL).
 FT NON_CONS 1010 1011 C4-TYPE (POTENTIAL).
 FT DOMAIN 2300 2325 POLY-PRO.
 FT DOMAIN 4391 4442 PDZ.
 FT DOMAIN 4544 4633 C2 DOMAIN 1.
 FT DOMAIN 5031 5121 C2 DOMAIN 2.
 FT VARSPLIC 4404 4404 QTGLMEG (in isoform 2).
 FT VARSPLIC 4534 4534 K -> KPTDGTKVVSHPITGEIQ (in isoform 2).
 FT VARSPLIC 4576 4576 G -> GOVMVVQNAS (in isoform 2).
 FT VARSPLIC 4757 4761 TAHKS -> SKRRK (in isoform 2).
 FT VARSPLIC 4762 5147 Missing (in isoform 2).
 FT VARSPLIC 5147 AA; 563537 MW; CD5D84990498CD3C CRC64;
 SQ SEQUENCE 5147 AA; 563537 MW; CD5D84990498CD3C CRC64;
 Query Match 2.9%; Score 421.5; DB 1; Length 5147;
 Best Local Similarity 18.1%; Pred. No. 1.8e-07;
 Matches 530; Conservative 390; Mismatches 940; Indels 1067; Gaps 128;
 QY 275 LECVKPPL--EEVP-----EDEVOCEVCVAHKVPGVTDVCAEATQKNKPYIRHPIGY 324
 DB 1076 LKVKETLSMEKIPPMVTVDQKQESKLE-----KDKASALQEKKPLPE-----1119
 QY 325 DRSRRKWFNLRLIIEETENENKKINWYSTYKVQLAEILDCLDNDYWEALCKLILEEM 384
 DB 1120 -----ERKLPEEKIRSEKK-----PLLEE-1141
 QY 385 REBIHRMDITDLTNKARGSNKSLAAANEELIESTRAKKGDIDNVKSPPEETEKQNE- 443
 DB 1142 -----KKPTPEDKLLPEAKTSAPEE---QKHDL--LKSOVQAEKLEG 1181
 QY 444 -----TENDSKDAEKNREEFED---QSLEKSDDKT-----PDDDPQEGK 480
 DB 1182 RVAPKTVQEGKOPQTKMEGLPSGTQSLPKE-DDKTTTKIKEQPQPPCTAKPQEKEDDK 1240
 QY 481 SEVGDFKSKNGELSESPCAGKAGSGSTRITRLRNPDLSKLSQLAKSOQVAAAHANKL 540
 DB 1241 SOTSSQQPKSPGSLDGTGYSSDGISSLSGEIPSLIPTDEK-DILKGLKKDSFSQSSP- 1298
 QY 541 FKEGKEVLVNVNQSGLSRSLSTKKEVIMKGNINNYFKLGOEGKYRVYH-----NOYSTN 593
 DB 1299 -----SSPSDLAKLESTVLSLEAQASTLADEKSEKTKQPHVSPQPKDQKQTK 1348
 QY 594 SFA-----LNKHQHRDHDKRRHLAHLFCLTPAGFEKWNKSVHSGK---VLITSLRLT 644
 DB 1349 SLSETLEITISEEIKESQBERKDTFKDSQQIPSSK-----DHKEKSEFVDDITTRREP 1404
 QY 645 ITQLENN-----IPSSFLHPNWASHRANWIKAV-QMCKSPREFALALAILCAVKPVV 696
 DB 1405 YDSVEESSESENPSVPQRRKRTSVGSSSDSEYKQDSQSGSEEDFIRKQIIEMSADEDA 1464
 QY 697 MLPINREFLGHRLHMTISEREKEKVKKKER-----729
 DB 1465 SGSEDEFTINQLKEISSSTESQKKEETKCKGKITAGKHRLTRKSTSTSIDEDAGRHSW 1524
 QY 730 KQEEETMQOATWVKYTFPVKHQVHKQKGEYRVTYGGWSWTSKTHYRVFKPLPNTN 789

```
Db 1525 HDEDEAFDESPELKY-----RETKSQESELYVTGGG-----LRRF----- 1562
Qy 790 VNYRKSLEGTGNMDENMDSDRKCRSPKKIKI-----EPDSEKDEVKGS-----DAAGAD 843
Db 1563 -----KTIENSTIADKYSAESQKTKSYLFDPEPELEMESLTDSPDRGRGEG 1611
Qy 844 QNEMDISKIPEKKDQDQKELLDSDKPKCKEPEMEYDDMDKTESHVN-----CBESSQV 897
Db 1612 SSSIHASFTPGTSPTSVSLDESDSPSHKKGESKQQRKARHPGHPLLPTIEDSSEE 1671
Qy 898 DVNVSEGFHLRTSYKKKTKSSKLDGLLERRIKQFTLEKQKORLEKIKLEGIGIKGTST 957
Db 1672 EELREEE-----ELLKQEQRQREIQQR-----KSSS 1699
Qy 958 NSKNLSSEPVITKAKEGCOSDSMROE-QSPNANNDQPEDLIQGCQSQSSVLRMSDPHS 1016
Db 1700 KSK-----KDKDELRAQRRRRPKTPPNLSPIED-----ASPEELRQAEME 1744
Qy 1017 TTNKLYPKRVDVDSIRSPTKPKON-SIENDIE-----BKVSDLASRGQEPKSKT 1069
Db 1745 ELHR-----SSCSYSPSIESDPEGFEISPEKIIIEVQVKYKLPTA--- 1784
Qy 1070 KGNDFFIDDSKLASADDIGLICKNKKPLIOESDTIVSSSKSALHSSVPKSTNDRDATP 1129
Db 1785 -----VSLYSPTD-----EQSIMQKESQKALKSAEMEYEMMHKTHYKAFP 1827
Qy 1130 LSRAMD-----FPGKLGCSSESNTLENSDVTISQDSEEDMIVQNSNESIEQFRTREQDV 1187
Db 1828 AANERDEVFE-----KEPLYGGMLIEDIYESLVEDTYNGSDVGSLLTROE-- 1873
Qy 1188 EVLEPLKCELVGSESTGNCEDRLPVKGTANGKKPSQKLEERPKNKSDQIKLKNWTD 1247
Db 1874 -----EENGFMQO-----KGRE---OKIRLSEQIYEDPMOKITDLOK----- 1907
Qy 1248 KKNNEPSEKKQORTSTFOINGKDNKPKIYLAGECLKEISESRVSGNVEPKVNNINKI 1307
Db 1908 -----EYELSHSVVPOEDIVSSSF-----I 1930
Qy 1308 IPEN-----DIKSUTVKESAIRPFINGDVIIMEDFNER-----NSSETKSHL-----LSS 1351
Db 1931 IPESHEIVDLGTWVTSFEERKLLDADAAYEELMKRQOMQLTPCSSPTQAPIGEDWTEST 1990
Qy 1352 SDAEGNVRDSL---ETLPTKESDSTQTT---PSASCPESNSVNQVEDMEIE---TSEVKK 1404
Db 1991 MDFRMPDASLTSVLSGASLTDSTSSATLSIPDKVITQHFSTEIEIEDEYVTDYTBRIQE 2050
Qy 1405 VTS---SPITSEESNLSDIDENGLPINKNENNGESKRTVITEVTTWTSTVAT--ES 1460
Db 2051 IIAHESLILYSESESATVPFSDTP-----SLTSSVSSVCTTDDSSPITTLDS 2100
Qy 1461 KTVIKVEKGDQTVSSSTENCAKSTV-----TTTTVTVKLSTPSTGGSYD--IIS 1509
Db 2101 ITTVYTEPDMITKFEDSEIESSSTYPGSIIDYPEEISVSLDRTAPPDGRASADHIVIS 2160
Qy 1510 VKEQSKTVVTTVTDSTTTGGTLVTSMTVSKYSTDKVKLKFSPKTRSG---TAL 1566
Db 2161 LSDMASSIISESVPKPEGPVADTVSTDLLISE-----KDPVKKAK-----KETNGIILLEV 2212
Qy 1567 PSVR---KFVTKSTKKS---IFVLPNDLKLARKGIREVYFNY-----NAKPALDI 1614
Db 2213 EAYRDKKLEBAERTKSLSETVFDHPSPSSVIALPMEQOL-STTYFTSGEFTGQEKPASQL 2271
Qy 1615 WPVPS-----PRPTFGITWRVRLQTVKSLAGVSLMLRLLLRWDDMAAKVPP 1663
Db 2272 ---PSGSPSVSSLPAPKPRPF-----RS-----SSL---DISAQPPP 2302
Qy 1664 GGSSTRTETSETITTTTTEILIKRRDVGPYIRFEYCIKIKICPIGVPTPKETPPQKRL 1723
Db 2303 -----PPPPPPPPPPPPPP----- 2316
Qy 1724 RSSALRPKRPTPKQTG--PVIIETWVAEBELEWEIRAEVEREKEKAQAEQOAKRL 1781
|||
```

```
Db 2317 -----PPPPPLPPPTSPKPTIL----- 2338
Qy 1782 EQOKPTVIAT-----STTSPTSTSTISTISPAQKVM-----VAPISGSVTTG----- 1822
Db 2339 TVASPVTTATPLTDAVTTLETTLAVLSRNGLPVTRICTTAPPPVPKPPSSIPSLGVTHRP 2398
Qy 1823 --TKMWLTTKVGSFPAIVTFOQNKNFHOTFATWVKQGSNSGVVQOVQKVLGIIPSTG-- 1878
Db 2399 EPSKPIAPKPVLPQPT-----TTQKPTDIHPKPTGLS 2432
Qy 1879 -TSQOIF-----TSFQPRATVITIRPNTSGSGTTSNS-----QVLTGQPIRPGMIVIRT 1927
Db 2433 LLSMTLNLVTSADYKLPSPTSPLSHSNKSSPRFSKSLTETVTVITLPS-EPG-----T 2486
Qy 1928 PLOQSTLGKAIIRTPVMVOPGAPQVMTQIIRQPV-----STAYSAP----- 1970
Db 2487 P-TDSSASQAITSMLP-----GSPSKDLVSV--EPVSVVPPVTAVEIPISSSQTEFVIG 2538
Qy 1971 --NTVSTPGQKSLTSATSTSNIOSSASQPPRQOQOVKLTMAQLTQLTQGHGNOGLTV 2028
Db 2539 ALQTFSATP-----VTAFSSFAAPTSTVQFLTTEVSKTEVSATRSTAPSGLSSISI 2591
Qy 2029 VI-----OGOGOTTGLOLIPQGVTVLPQP----- 2053
Db 2592 TIPEPLALDNIHLEKPKYKEDGKLQVLGVDIRVTPKVEVKTDKCIDLSASTMDVKR 2651
Qy 2054 -----GOOL--MQAAMPNGTVQRPFLTP--LAT-TAT-----TASTTTTTVSTTAAGT 2096
Db 2652 QITANEVYGKQISAVQPSIINLSVTSSIVTPSLATETVTFVTCTASASTTESILVGA 2711
Qy 2097 GEOROSKL-----SPQMOVHQDKTLPPA-OSSSVGPAKAPQAPQAPQAPQOTOP- 2146
Db 2712 EHAMTTPLOLTKSHAEPYRIPSDQVPIAREEAPINLSGLTPAHAVTLAITKPVTPVP 2771
Qy 2147 -----QSPAQPEVQTOPEVQTOVTVSSHVPSEAQPTHAQSSKPOVAQAQSQSQSVQ 2198
Db 2772 VGTNGWTDVTSQGI--TDEGVVDLSTTKSHRIVTMDSTSSVMKLIIEDEKPDVLTAG 2830
Qy 2199 QSPV-----RVQSPSQT----- 2215
Db 2831 RRAVCCDVVKLPFGRSCTAQAPATLTPEDRFYGRDDHYQYDRSGPYGYRGIGMKPMS 2890
Qy 2216 -----TPQSLSPGQSQVQVQTTTSQPIQPIPHSLQ 2245
Db 2891 DTNLAEAGHFFYKSNADYSEGTDRAVLTSGRVTGGEVMDYSSKTTGYP-----ETQ 2946
Qy 2246 IPSQGOQSQSQVQSQSTQLSSG--QPLNOVSVSSSRPQLQIQOQPOQVIAVPQ----- 2298
Db 2947 VIS-GAGISTPQYSTARMTPPGPOYCVGSLRSSNGVVYSSVATPTPTFAITTPQPSI 3005
Qy 2299 LOQOVQVLSIQ-SOVVQAOIAQOQSGVPQO-----IKQLPIQIQS 2339
Db 3006 FSTTVRDLSGIHTADAVTSLPAMHHSQPMPSRYFITTGASSETDIANTGIDISASLQITM 3065
Qy 2340 SAVQTHOIQNVIVVQAASVOEQLQVQOOLRQO-----OQKKQOQIEIKREHTLQASN 2392
Db 3066 ESUTAETIDSVPLTTAS--EVFPEV--VGDESALLIVPEEDKOQOQDLDER----- 3113
Qy 2393 QSEIIQOVVMKHNNAVIEHLKQKSMTPAEREENQRMIVCNQVMKYITLIDKREEKQAAK 2452
Db 3114 --ELLE-----LEKIKQOR-----FABELEWERQBIOR 3139
Qy 2453 KRKREESVEQKRQONATKLSALLFKHKQLRAE-ILKKRALLDKDLOIE-----VOBELK 2507
Db 3140 FREQEKIMVQKLEELQSMKOHLLFQOEEREQAQFMQRQETLAQOQLQEQIQOQLQOOLH 3199
Qy 2508 RDLKIKEKDLML-----AQATA---VA 2528
Db 3200 QOLEEQKIRIQYNYDPDPSQTTEQAILEGOVAALEGQFWATEDATTASAWVA 3259
Qy 2529 APCPP-----VTPVLPAAPPAPPPPPPPGVQHTGLLST-----PTLPVASOKRK 2573
Db 3260 IEIPQSGWTVQSDGVQYI---APP-----GILSTVSEIPLTDVVVREK 3303
```

```

QY 2574 REERDSSSKKKKMMISTTSKTKKDKLYCICKTPYDESKFYIGCDRCQWYHRCVG 2633
D 3304 QPKRSSGAKVRGQ-----YDDMGENTDDP----- 3329
QY 2634 ILOSEALIDEVQCOSTEDAMTVLPLETKDEYGLKR-----VLSLQAHKAWPFL 2688
D 3330 --RSPFKIVDSGV-----QTDD-----EDATDRSVSRRTTKKSVDTSVQ----- 3368
QY 2689 EPVDPNDAPDYGVIKEMDLATMEERVQR--RYEKLTEFEVADMTK 2733
D 3369 --TDDEDQEW-----DMPTRSRKARVKYGDMSMT--ADTK 3403

RESULT 6
BA2A_MOUSE STANDARD; PRT; 1850 AA.
AC Q91YE5.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Bromodomain adjacent to zinc finger domain 2A (Transcription
DE termination factor-I interacting protein 5) (TFI-I interacting protein
DE 5) (Tip5)
GN BA2A OR TIP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423468; PubMed=11532953;
RA Strohnner R., Nemeth A., Jansa P., Hoffmann-Rohrer U., Santoro R.,
RA Laengst G., Grumet I.;
RT "NORC-a novel member of mammalian ISWI-containing chromatin remodeling
RT machines";
RL EMBO J. 20:4892-4900(2001).
CC -!- FUNCTION: May play a role in transcriptional regulation
CC interacting with ISWI. May serve a specific role in maintaining or
CC altering the chromatin structure of the rDNA locus.
CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NORC
CC (nucleolar remodeling complex).
CC -!- SUBCELLULAR LOCATION: Nuclear. Co-localizes with the basal RNA
CC polymerase I transcription factor UBF in the nucleolus.
CC -!- SIMILARITY: BELONGS TO THE WAL FAMILY.
CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
CC -!- SIMILARITY: Contains 1 methyl-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ309544; CAC69992.1; -
CC MGD; MGI:2151152; Baz2a.
CC DR GO:0005731; C:nucleolus organizer complex; ISS.
CC DR GO:0003677; F:DNA binding activity; ISS.
CC DR GO:00030528; F:transcription regulator activity; ISS.
CC DR GO:0006338; P:chromatin modeling; ISS.
CC DR GO:0006335; P:regulation of transcription, DNA-dependent; ISS.
CC InterPro; IPR000637; AT_hook.
CC InterPro; IPR001487; Bromodomain.
CC InterPro; IPR004022; DDT_dom.
CC InterPro; IPR001739; Methyl-CpG_bind.
CC InterPro; IPR001965; Znf_PHD.
CC Pfam; PF02178; AT_hook; 4.
CC Pfam; PF00439; bromodomain; 1.

```

```

DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 4.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS0014; BROMODOMAIN_2; 1.
DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;
KW Nuclear protein; Repeat; DNA-binding.
FT DOMAIN 505 557 MBD.
FT DNA_BIND 602 614 A.T HOOK 1.
FT DNA_BIND 623 635 A.T HOOK 2.
FT DOMAIN 800 862 DDT.
FT DNA_BIND 1137 1149 A.T HOOK 3.
FT DNA_BIND 1351 1363 A.T HOOK 4.
FT ZN_FING 1623 1673 PHD-TYPE.
FT DOMAIN 1755 1825 BROMODOMAIN.
FT DOMAIN 613 738 LYS-RICH.
FT DOMAIN 647 774 COILED COIL (POTENTIAL).
FT DOMAIN 1006 1012 POLY-GLU.
SQ SEQUENCE 1850 AA; 205597 MW; E5DD4FED1D46DAE1 CRC64;

Query Match 2.9%; Score 419.5; DB 1; Length 1850;
Best Local Similarity 18.0%; Pred. No. 6.6e-08;
Matches 386; Conservative 295; Mismatches 776; Indels 693; Gaps 81;

QY 956 STNSSKNLSSESPVI-----TKAEGCQSDSMRQEQSPNA 989
D 61 SANPNLKDPPLLSQFPGQYPLNGILGGRQSPSPSHNTNLRAGSQEFWANGTQSPMG 120
QY 990 NNDQPEDLIQCSQSDSSVLRRMSPDSSHNTNKLKDRVLDVDSIRSPKCPKQSIEND 1049
D 121 LNFDSQELYSPPDQNFVMPNGPPSPFTS---PQTSPLMGSSI---QTFAPSQD-VSSD 173
QY 1050 I-----EEKVSDLASRGOEPTKSKTGNDFIDDSKLASADDITGLCKKKKPLIOEE 1102
D 174 IHPDRAAEKELTSVVAENG-----TGLVGSLEEEQELKMC 211
QY 1103 SDTISSSSKSLHSSVPKSTNDRDATPLSRAMDFEGKLGCDSESNSTLSSDPTVSIQDS 1162
D 212 GYNGSVSSVESLHQEVSVLPD---PTVSCLD-----DPSHLPDQ 248
QY 1163 SEEDMIVONSESISQFRTREQDVEVLEPLKCELVSGESTG-----NCDRLPVKGT 1215
D 249 LEDTPILS-----EDSLEPFDLSAAEPVSGSLYIGIDDAELMGAEDKLPLEG- 294
QY 1216 EANGKKPSQOKLEBPPVKNKCSQDIKLKNTTDDKKNENRESEKKGORTSTFOINGKNKP 1275
D 295 -----NPVISALD----- 302
QY 1276 KIYLAGCELIKESRVSNGVNEPKVNNIN--KIIPENDIKSLTVKESAIRPFINGDVIM 1333
D 303 -----C-----PALSANAFSLADDQSQTSAISFVSPSPVLPGESVL 340
QY 1334 ED--FERNSSSTKSHLLSSDAEGNYRDSLTLPSTKESDSTQ-----TTTTASACP 1384
D 341 QDNSEGLNSCSDSEQEIEIETQSS--NFQRP-L-TEPAPQPPSTQLHPAVSPASPAASLT 397
QY 1385 ESNYSNOVEDMEIETSEVKKVTSSPTSEESNLSNDFIDENGLIPINKNENVGESKRKT 1444
D 398 ASAEISPA-----VSPVASSVPV----- 415
QY 1445 VITEVTTMTSTVATESKTVIKVEKDGKQTVVSSSTENCARKSTVTTTTTKLSTPSTGGS 1504
D 416 --PEVFAVSPASSPALPAISLE-----ASMTPTVSPQSGSPSPAAA 457
QY 1505 VDIISVKEQSKTVVTTTVDTSLTGGLTVTSMT---VSKEYSTRDKVKL---MKFSRPK 1558

```


QY 1485 TVTTTTTTLKSTPSTGGSDIISVKEQSKTVVTTTVDLSLTGTTGLTSMVTSKEYS 1544
D 1486 PASSKRLKTSLG--KVGVEKELLAVGLKTQTSGTTHHTHTPTGDKSMKAPMSPKQI 1683
QY 1545 TRQVKVLMKSPKPKTRSGTA-----LPSYRK--FVTKSTKKSIFVLPND 1588
D 1684 LDSAASLITGSKRLRTPKSKSEYEDLAGFIELFQTPSHTKESMTNEKTKVSYRASQPD 1743
QY 1589 L-----KKLARKGGIRVPVFNNAKPAALDIWVPSRPFTGIIWRYRLQTVKS 1637
D 1744 LVDTPTSSKQPKARSLRKADTEE-----EFLAFKQTPSAG-----KAMHTPKP 1787
QY 1638 LAGVSLMLR--LLWASLRWDDMAAKVPPGGGSTRTETSEITTEITLTKRDVGPYGR-- 1694
D 1788 AVGEKINDITFLGTPVQKLDQPCNLN--GSNRLQTRKEKAQALELT-----GFREL 1838
QY 1695 -----FEYCKRKIC--PIGVP--ETPKETPTPQKGLRSS-----ALRPKRP 1733
D 1839 FQTPCTDNPTADEKTTKKILCKSPQSPADPTPTNTPQPKRSLLKADVEEFLAFKLTP 1898
QY 1734 -----ETPKOTGPVLIETWAAEELWEIRAF--AERVE-----KE 1768
D 1899 SAKAMHTPRAA-----VGEEK-----DINTFVQTPVEKLDLNLNPGSKRRPQTPKE 1946
QY 1769 KAQAEEQAKRLEQOKPTVIATSTTSPTSTISPAQKVMVAPISGSVTTGTWKVLT 1828
D 1947 KAKALEDLAGKELFQTPGHTSEEMTDKITEVSCSKSPQDPVKTP-----TSSKQRLK 2000
QY 1829 TKVGSPATVFOONKFNHQTATVWQGGNSGVQVQKVL--GIIPSTGTSQQTFFS 1886
D 2001 ISLGK-----VGVEEVLPVGLTQTSKTTQT-- 2028
QY 1887 FQPRATVTVIRPNTSGSGP-----TSNSQVI-----TGPOIRPGMTVIRTP-----L 1929
D 2029 -----HRTAGDKSIKAFKESAKQMLDPANYGGMERWP-----RTPKERAQSL 2073
QY 1930 QSTLTKKAIIRTPVMPQGAPOQVMTQIIRGPQVSTAVSAPNTVSTSPGQKTSATS 1989
D 2074 EDLAGKELFQTPDHEESTTDDTKTKI-----ACKSPPESEMDTPTST-- 2117
QY 1990 NIOSSASQPRPOGQVKLT--MAQLTOLTQGHGNOGLTVVLOGOQTGQLQLIP-- 2044
D 2118 -----RRRKTPLGKRDIVEELSALKQLTQ-----THTTDDKVPGE 2153
QY 2045 -QGVTVLPGGQQLM-----QAAMPNGTVQ-----RELFTPLATTATTAS 2083
D 2154 DKGINVFRETAKQKLDPAASVTSKQRPRTPKGAQPLEDLAGLKELFQTPVCTDKPPTH 2213
QY 2084 TTTTTTSTTA-----AGTGE--QRQSKLS--PQMVHODKTLPPAQSSSVGPAKAQOTA- 2134
D 2214 EKTTKIACRSPQDPVGTPTIFRQSKRSRLRKADVEESLALRKRTSPVGKAMDTPKPA 2273
QY 2135 -----QPSARPOQTQPSQAQPEVQTPQVQTTVSS-----HVP 2171
D 2274 GDEKMAKFMGTVPVKLDLPGNLPGSKRWQTPKE-----KAQALEDLAGKELFQTP 2326
QY 2172 SEAPHTAQSCKPQVAAQSPQSNVQSGSPVRVQSPQSTRIPS----- 2215
D 2327 GTDKPT--TDEKTKIACKS--PQ-----PDPVDVTFASTQKRPKRLNKADVEEFLALR 2377
QY 2216 -----QLSPGQSQVQTTTSQPIPIQPHSTL-QTP-SOGQPSQSP----- 2256
D 2378 KRTPSAGKAMDTPKPAVSDEKNINTFVETPVQKLDLNLNPGSKRPQTPKERALEDL 2437
QY 2257 -----QVSSSTQLSSGGTFLQVSVSSPSRQLOIQOPQVIAVP-----QLQQQVQV 2305
D 2438 VGRKELFQTPGHTSEEMTDKITEVSCSKSPQSPESFTKTSRSKORLKIPLVKVDMKEPLA 2497
QY 2306 LSQT--OSQVVAQIQAOQSGVPOQIKL--QLPIQIQOSSAVQTHQINQVVT--VQAASVQE 2360
D 2498 VSKLTRTSGTQTHTPTGDSKSIKAFKESPKQILDPAASVTSGRRLTRKEKARALE 2557

QY 2361 QLQRVQOL-----RQOQK-----KQOQIEIKR--- 2384
D 2558 LDVDFELFSAPGHTSEEMTDKNTKIPCKSPPELDTATSTKCPKTRPRKEVEELS 2617
QY 2385 --EHTLQAQNSOELIOQVVMKHNAVIEHLKQ--KKSMTPAAREENORMIVCNOVMKYIL 2440
D 2618 AVERLTQTSQSTHTHKEPASGDEG--IKVLQKAKKPNPVEEPSRR----- 2664
QY 2441 DKIDKEEKAARKKREESVEQKRSKQNAATKISALLFKHKEQLRA-----EI 2487
D 2665 -----RPRAPKEKAQPLEDLAGFTELSETSGHTQESLTAGKATKIPCESPPLEV 2713
QY 2488 LKRRALLDQDLQIEVQBELKRDLIKKEKDLMLQAO----- 2523
D 2714 VDTASTKRLTRVQK-----VQVKEPSAVKFTQTSGETTDDADKEPAGEDKIGIKALKE 2768
QY 2524 -ATAVAACPVPVPLPAPAPPSP-----PPGVQVHTGLLSTPLPVASQKRKR 2574
D 2769 SAKQTPAASVTVGSRRRPRAPRESAQAIEDLAGFKDPAAGHTEESMTDDTKTKIPCKSS 2828
QY 2575 EEEKSSSSKSKKKKMTSTTSKETKDKTKLYCICKTPYDESKFYIGCDRCQNMVHGRCVGI 2634
D 2829 PELEDATSSKRRPRTRAKQVKEB--LLAVGK-----L 2861
QY 2635 LQSEAELEIDYVQPCQOSTEDAMTVLPTKEKYEGKRLVRLSLQAHKAWPPEPDPN 2694
D 2862 TQTSGE-----THTDKEPVGEGKGTKA-----FKQAKRN 2892
QY 2695 -DAPDIYGVYKPEM-----DLATMEVRQR--RYEKLTFEFVAD 2730
D 2893 VDAEDVIGSRQRPAPKEKAQPLEDLASFQELSQTGCHTEELANGAAD 2940
RESULT 8
AMVH_YEAST STANDARD; PRT; 1367 AA.
ID P08640; P08068;
AC 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucon 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN STAL OR STA2 OR MAL5 OR YI019C.
OC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX PubMed=9169870;
RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D., Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N., Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G., Moulle S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N., Skellerton J., Walsh V., Whitehead S., Whitehead S., Barrell B.G.;
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX";
RL Nature 387:84-87(1997).
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RA "Gene fusion is a possible mechanism underlying the evolution of STAL";
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RA "Similar short elements in the 5' regions of the STA2 and SGA genes from Saccharomyces cerevisiae";
RT FEBS Lett. 239:179-184(1988).
RL

SEQUENCE FROM N.A.
 Yoon H., Philp N.J.;
 "Cloning of a new extracellular matrix protein expressed in retina.";
 Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May play a role in transcriptional regulation.
 CC -1- SIMILARITY: BELONGS TO THE WAL FAMILY.
 CC -1- SIMILARITY: Contains 1 bromodomain.
 CC -1- SIMILARITY: Contains 1 DDT domain.
 CC -1- SIMILARITY: Contains 1 methyl-binding (MBD) domain.
 CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF224275; AAG36791.1; -;
 DR HSP; Q92831; 1891.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR004022; DDT_dom.
 DR InterPro; IPR001739; Methyl-CpG_bind.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF02791; DDT; 1.
 DR Pfam; PF01429; MBD; 1.
 DR Pfam; PF00628; PHD; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00571; DDT; 1.
 DR SMART; SM00249; PHD; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
 DR PROSITE; PS00016; ZF_PHD_2; 1.
 KW Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;
 KW Nuclear protein; DNA-binding.
 FT DOMAIN 693 743 MBD.
 FT ZN_FING 1010 1072 DDT.
 FT DOMAIN 1895 1945 PHD-TYPE.
 FT DOMAIN 2039 2109 BROMODOMAIN.
 FT DOMAIN 9 46 SER-RICH.
 FT DOMAIN 184 208 SER-RICH.
 FT DOMAIN 212 219 POLY-GLU.
 FT DOMAIN 544 610 ASP-RICH.
 FT DOMAIN 797 984 COILED COIL (POTENTIAL).
 FT DOMAIN 882 886 POLY-LYS.
 FT DOMAIN 1219 1262 ASP-RICH.
 FT DOMAIN 1909 1912 POLY-LEU.
 SQ SEQUENCE 2130 AA; 236140 MW; 208C48FB0BA68F70 CRC64;
 Query Match 2.8%; Score 396.5; DB 1; Length 2130;
 Best Local Similarity 17.9%; Pred. No. 5e-07;
 Matches 464; Conservative 345; Mismatches 902; Indels 875; Gaps 104;
 QY 419 ESIRAKGDIDNVKSPETEDKNE-----TENDSKDAEKNEEFEDQSLKSDSD 469
 DB 170 DKIKDKPRKAVESSNSDSGSSDTSSEGISSDSDLEDEDEEEEDQSAESED 229
 QY 470 KTPDDDP--EOKSEV---GDFKSEKNGELSESPGAGKAGSGSTRITRLRNPDKLSQ 524
 DB 230 ESDSENEAHENKNKVLHSGVKDMKTDGKAHEKSQEKTRHQIPLVSDSQTHSFQSQ 289
 QY 525 LKSOQVAAA---AHEANKLFE--GKEVLVNSQGEISRL-----STKKEVIMK--- 568
 DB 290 QKQPVLSQQLPFFIQSSQAKEESVKNKTSITQSTGLVPNVKPLSLVHQTAKKAYLKIIV 349
 QY 569 -----GNINNVFKLQGEQKRYRYHNQYNTSNFALNKHQHREDHDKRRHLAKFCLTP 620
 DB 350 PPPDLLKAGNKT-----SEESIPLIS-----DVRSKREQYKQTF---P 385

QY 621 AGEFKWN-----GSVHGSKVLTTSTLTTLTITQLENNIPSSFL-HPNASHRAWNI 669
 DB 386 AALQKQESSKNLKKVIAISLSSKPTSCSPAQKLTSLNNHSPFLTNALIGHPNGV 445
 QY 670 KAVQMSKPREFALAILCAVAVVMLPIWREFLGHTRLRHMTSIEEREKVKKEK 729
 DB 446 -----IOSVIOEVPLALTUKOK 462
 QY 730 KOEE-EETMQATWVKYTPFKVHKQVKGEEYRVTVGGWSWIKTHVYRFV-PKLPON 787
 DB 463 SQTINESVAIASSTPFSPLPVNLSAGKK-----FTG-----NRTLVPSTPVLPGS 510
 QY 788 TNYNRSKLEGTNNNDENDESDDRKCRSPKKIKIEP-----DSKDEVKGSDDAKGAD 843
 DB 511 -----GKDKPVSNNAVAVKTHCLPSAKLVLEQVGRGVSDAPSKESDSDSDDD 560
 QY 844 QNEMDISKITEKDDQVKKELLDSDSKPCKEPEMEVDDDMKTESHVNCQESSQDVVNV 903
 DB 561 DDEDEDEDEDDSDSQ--SESDSNSESDDGSEDEDEDDK-----QDESDD---T 611
 QY 904 EGFHLRTSYK-KTKSSKLDGLLERRIKOFTLEEKORLEKIKLEGGIKGIGKTSTNSKN 962
 DB 612 EG--EXTPLKLTGSS-----IKS-----SSIGPVAHSTPLN 642
 QY 963 LSESPVITRAKEGCSQSDMRQEQSPNANDQEDLLTQCSQSDSVLRMSDPSHTNK-- 1020
 DB 643 L-----QVAKTPSSAPAL---CPETQPAVFLGTTSTPLPSSH 678
 QY 1021 --LYPKDRVLDVSIKSP-ETKCPKNSIEN---DIEKVSOLASRGQEPKTSKTKGDNF 1074
 DB 679 CGISKRRVYDERELRVPLEYGHQRETRIRNFGRLQGEVAYAPCGK----- 726
 QY 1075 FIDDSKLASADDI--GTLLCKNNKPLIQEESDTIVSSSKSALHSSVPKSTNDRDAPLSR 1132
 DB 727 ----KLQVPEVVKGVQWC-----LLKEE-----EVVPCIR 753
 QY 1133 AMDFEKLGCDSNSTLENSDTSVSIQSDSEEDMIVQNSNESISQFRTQDV---EV 1189
 DB 754 AM--EGRRG--RPPNPDQHSRE-----ESMRRRGRPPNVGSTF 791
 QY 1190 LEPLKCELVSGESTGNCEDRLPVKGTANGKPKSOQKLEERPKNKCSQDIKLNTDOK 1249
 DB 792 LDSTDAKLL-----RKLQAEIARQAQIKLLRLKQK 824
 QY 1250 NNNRSEKKGQRTFQINGKDNKPKYILKGLCEISESRVSVGNVEKVNINKIIP 1309
 DB 825 EQARAAKEAKQQA---IMAAEEKR--QKQIKIMKO-----QEKIKRIQOIRM 869
 QY 1310 ENDIKSLTVKESAIRPFGDVIWEDFNERNSETKSHLLSSSDAEGNYRDSLETLPSTK 1369
 DB 870 EKELRA-----QQILEAKKKKEAANAALL--EAEKRIKE-----K 904
 QY 1370 ESDSTQTTTPSASPESNSYNQVEDMEIETSEVKVYTSPTSEESNLSNDFIDENGLP 1429
 DB 905 EMRRQAVLLKHQELERHRLDMERERRRQHMMLKAMEARKKAEEKERLQKREDEKRL- 963
 QY 1430 INKNENNGSKRKTIVITEVTTTSTVA--TESKTVIKVEK-----GKQTVSS 1477
 DB 964 -NKRLEQRRLLEMAKELKKNEDMCLADQALPELPRIGLVLSGTSFSCMLIVQF 1022
 QY 1478 TENCASSTVTTTTTTLKSTPSTG-----GSVDIIISVKEQSKTVTTTVDLSLTT 1529
 DB 1023 LRNFQKVLGPDVNTDVSLSLTQELGLNIGDSRGEVDLLVK-----LVTAACVDCPLVT 1077
 QY 1530 GGTLLVTSMTYSKEY-----STRDKVKLM-----KFSRPKTRSGTALPSYRKF 1572
 DB 1078 G---YKAKTILGEHLLNAGVNRDNVSEILQIFMEAHGQOTELTESLTKAFQA----- 1127
 QY 1573 VTKSTKKSIFVLPNDLLKKLARKGIREV-PYFNNAKPAIDTWIPSPRPTFGITWYR 1631
 DB 1128 HTPACKAAVLAFLVNEL--ACSKSVSWSEIDKNIDYMSNLRDKW----- 1169
 QY 1632 LQTVKSLAGVSLMLRLWLASLRWDDMAKVPVPGGSGSTRTETSETETITTEIKKRDVGPY 1691


```

Db 3909 TPPTPTGTGTGTPTTTTPIITTTTTP-TPPTGTGQ-----TPPTTPIITTTTTPPTPTGT 3963
QY 1983 TSATSTSNIOSSQP-PRPQQGVKLTAQLTQLAQGHGNOGLVVIQGGQGTTCQLQ 2041
Db 3964 QTPTTPTTITTTTTPPTPTGTGTGTPTTPTTITTTT-----VPTPTPTGTQPTTT 4015
QY 2042 LIPQGVTVLPQGVQLMQAMPNGTVQRELFPTPLAT-----ATTASTTT 2086
Db 4016 PTTTTTTPPTP-----TPPTGT-QTPTTPTTITTTTTPPTPTGTGTPTTPTTITTT 4066
QY 2087 TVVSTTAAGTGEOQRQSKLSPQMVDKTLPPAQSSVGVGPAQAOPATOPASRPQPTOP 2146
Db 4067 TVTPTPTPTGTGTPT-----TPPTTITTTTTPPTPTGTGTPTTPTTITTTT 4114
QY 2147 QSPAQEVQTPQVQVQTTSVSHVSEAPQTHAQSSKPVAAQSQSNVQGSQSPVRVQS 2206
Db 4115 VTPTPTGTGTGTPTTPTTITTTTTPPTPTGTGT-PTTPTTITTTTTPPTPTGTGT 4172
QY 2207 PSQTRIRPS---TPSQLSQGQSQVQVTTTSQPIQPHITSLOIPSGQGPQ-----SQPQV 2258
Db 4173 PTTPTTITTTTTPPTPTGTGTGTGTGTPTTSTAPIAELTTSNPPPSSTPTQTSRSTSP 4232
QY 2259 QSSTQTLSSGQTLNQVSVSPSRP 2282
Db 4233 ESTTLLSTUPPAIEMTSTAPPSTP 4256

RESULT 11
PCLO_MOUSE STANDARD; PRT; 5038 AA.
AC Q9QYX7; Q9QYX6; Q9QZJ0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Piccolo protein (Presynaptic cytomatrix protein) (Aczonin) (Brain-
GN PCLO OR ACZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, ALTERNATIVE SPLICING,
RP TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.
RC TISSUE=Brain;
RX MEDLINE=99439764; PubMed=10508862;
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA Killmann M.W.;
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin.";
RL J. Cell Biol. 147:151-162(1999).
RN [2]
RP REVISIONS.
RC TISSUE=Brain;
RA Killmann M.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 4502-4682 FROM N.A.
RC TISSUE=Brain;
RA Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking.
CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin.
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9QYX7-1; Sequence=Displayed;
CC Name=2;

```

```

CC -!- IsoId=Q9QYX7-2; Sequence=VSP_003928, VSP_003929;
CC TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in
CC stomach. Not detected in other tissues analyzed including adrenal
CC gland, testis and pancreas.
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; Y19185; CAB60731.2; -.
CC EMBL; Y19186; CAB60732.2; -.
CC EMBL; AF181269; AAD55786.2; -.
CC HSSP; P04110; 1A25.
CC MGD; MGI:1349390; Pclo.
CC GO; GO:0045202; C:synaptic junction; IDA.
CC GO; GO:0005509; F:calcium ion binding activity; ISS.
CC GO; GO:0005544; F:calcium-dependent phospholipid binding acti. .; ISS.
CC GO; GO:0005522; F:profilin binding activity; IDA.
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; IDA.
CC GO; GO:0016080; P:cytoskeletal vesicle targeting; NAS.
CC InterPro; IPR000008; C2.
CC InterPro; IPR001478; PDZ.
CC Pfam; PF00168; C2; 2.
CC SMART; PF00595; PDZ; 1.
CC SMART; SM00239; C2; 2.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS00499; C2_DOMAIN_1; 1.
CC PROSITE; PS00004; C2_DOMAIN_2; 2.
CC PROSITE; PS0106; PDZ; 1.
CC Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW Repeat; Alternative splicing.
FT DOMAIN 371 470
FT 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT P-A-K-P-Q-P-Q-P-X.
FT ZN_FING 502 526
FT C4-TYPE (POTENTIAL).
FT ZN_FING 967 990
FT C4-TYPE (POTENTIAL).
FT DOMAIN 2305 2329
FT POLY-PRO.
FT DOMAIN 4394 4488
FT PDZ.
FT DOMAIN 4607 4705
FT C2 DOMAIN 1.
FT DOMAIN 4922 5012
FT C2 DOMAIN 2.
FT VARSPPLIC 4829 4833
FT TRPTN -> SKRRK (in isoform 2).
FT Missing (in isoform 2).
FT /FTID=VSP_003928.
FT /FTID=VSP_003929.
FT VARSPPLIC 4834 5038
FT Missing (in isoform 2).
FT /FTID=VSP_003928.
FT /FTID=VSP_003929.
SQ SEQUENCE 5038 AA; 547600 MW; DADA460CF3B40888 CRC64;

Query Match 2.7%; Score 390.5; DB 1; Length 5038;
Best Local Similarity 17.8%; Pred. No. 2.1e-06;
Matches 431; Conservative 330; Mismatches 885; Indels 773; Gaps 91;

QY 717 EREEKEKVKKKKKQ-----EEETMQQATWVYTPVVKHGVKQGEYRVYGYG 768
Db 1120 EEKALPADKKKKPPAAEAAPLEKKRPDP-----DQKLPDPAKPSASEGEKKRD----- 1170
QY 769 WSWISKTHYRVFPK-----LPNGTNNVYRK-----LPNGTNNVYRK----- 794
Db 1171 ---LLAHVQ--IPEESPIGVASLACEGQQPDPTRPDLPGATPQTLPKDRKQSRDVT 1225
QY 795 --SLEGT-----KNNMDENWDSKRCRSPPKKI----- 823
Db 1226 PQQAEGTAKGEGPSKDRTEKEEDKSD-TSSQQPKSPQGLSDTGYSSDGISGSLGEIP 1284
QY 824 -IEPDSEKDEVKG----SDAAKADQNMEDISKI-----TEKKDQD 859
Db 1285 SLIPSEKDLKGLKKKSFSSQESSPSPSLAKLESTVLSILEAQASTLVGEKAERKTP 1344

```



```

Db 3299 TDVVVKEEQPKRRSSGAK 3317
RESULT 12
PCLO_RAT
ID PCLO_RAT STANDARD; PRT; 5085 AA.
AC Q9JKS6; Q9JLT1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN PCLO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX MEDLINE=20170257; PubMed=10707984;
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a presynaptic zinc finger protein structurally related to
RT bassoon.";
RL Neuron 25:203-214(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
RP ALA-4694.
RX MEDLINE=2111819; PubMed=11285225;
RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
RT "An unusual C(2)-domain in the active-zone protein piccolo:
RT implications for Ca(2+) regulation of neurotransmitter release.";
RL EMBO J. 20:1605-1619(2001).
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin.
CC -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
CC junctions.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId-Q9JKS6-1; Sequence=Displayed;
CC Name=2;
CC IsoId-Q9JKS6-2; Sequence-VSP_003930, VSP_003931;
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF138789; AAF07822.2; -
CC EMBL; AF227534; AAF63196.1; -
CC HSSP; P04410; 1A25.
DR GO: 00045202; C:synaptic junction; IDA.
DR GO: 0005509; F:calcium ion binding activity; IDA.
DR GO: 0005544; F:calcium-dependent phospholipid binding acti...; IDA.
DR GO: 0005522; F:profilin binding activity; ISS.
DR GO: 0007010; P:cytoskeleton organization and biogenesis; ISS.
DR GO: 0016080; P:synaptic vesicle targeting; NAS.

```

```

DR InterPro; IPR000008; C2.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00239; C2; 2.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00499; C2_DOMAIN_1; 1.
DR PROSITE; PS00004; C2_DOMAIN_2; 2.
DR PROSITE; PS50106; PDZ; 1.
KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW Repeat; Alternative splicing.
FT DOMAIN 372 491 12 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT P-A-K-P-Q-Q-P-X.
FT ZN_FING 523 547 C4-TYPE (POTENTIAL).
FT ZN_FING 1010 1033 C4-TYPE (POTENTIAL).
FT DOMAIN 2351 2362 POLY-PRO.
FT DOMAIN 4442 4536 PDZ.
FT DOMAIN 4653 4732 C2 DOMAIN 1.
FT DOMAIN 4653 4732 C2 DOMAIN 2.
FT VARSPLIC 4876 4880 Missing (in isoform 2).
FT VARSPLIC 4881 5085 /FTID=VSP_003930.
FT VARSPLIC 4881 5085 /FTID=VSP_003931.
FT MUTAGEN 4668 4668 D->A: COMPLETE LOSS OF CALCIUM-BINDING
FT AND CALCIUM-DEPENDENT PHOSPHOLIPID
FT BINDING ACTIVITY.
FT MUTAGEN 4674 4674 D->A: COMPLETE LOSS OF CALCIUM-BINDING
FT AND CALCIUM-DEPENDENT PHOSPHOLIPID
FT BINDING ACTIVITY.
FT MUTAGEN 4688 4688 V->S: SMALL INCREASE IN AFFINITY FOR
FT CALCIUM.
FT MUTAGEN 4688 4689 VM->SS: 10-FOLD INCREASE IN AFFINITY FOR
FT CALCIUM.
FT MUTAGEN 4689 4689 M->S: INCREASED AFFINITY FOR CALCIUM.
FT MUTAGEN 4690 4691 VM->SS: 10-FOLD INCREASE IN AFFINITY FOR
FT CALCIUM.
FT MUTAGEN 4692 4693 QN->AA: MODERATE INCREASE IN AFFINITY FOR
FT CALCIUM.
FT MUTAGEN 4694 4694 A->S: NO EFFECT ON CALCIUM-BINDING
FT ACTIVITY.
SQ SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;
Query Match 2.7%; Score 388; DB 1; Length 5085;
Best Local Similarity 17.8%; Pred. No. 2.7e-06;
Matches 438; Conservative 334; Mismatches 832; Indels 860; Gaps 95;
QY 717 EREEKEKVKKKEKQEEEEE---TMOQATWVYKTPVVKHQVVKQKGEYRVYTCGWSWI 772
DB 1163 EEEKAISADKKERKPPAEKPPLEKKPIPVDKLPPEAKPLSSEGEKHE-----I 1214
QY 773 SKTHVYRVFPK-----LPGNTNVNYRKS-----LP 797
DB 1215 LKAHVQ--IPEEPTGKVAAKAGEEQPDSPREPALPGATPLTLPKAGEKERAQAQPAE 1272
QY 798 GT-----KNNMDENMDESKRCSRPKKI-----KTEPDS 828
DB 1273 GSSKDGQGBRSRSEKTEKEDKSDTSSSQPKSPQGLSDTGYSDDGSGSLGETPSLIPSD 1332
QY 829 EKDEVKG---SDAAKGADONEMDLSKITEKDDQVKELLDSDSD-----KPKCK 873
DB 1333 EKDLGLKGLKKDSFQESSSPSOLAHL-----ESTVLSTLEAQASTLVGEAKKTPQOK 1388
QY 874 EEPMEVDDDMKTESHVNCQESSQVDVNVVNSGPHLRYSYKKTKSKSLDGLLERRIKQFT 933
DB 1389 ISPEKPDQOKT-----QTASETLDTITSEE-EIKESQEKV-SPKKDS-----EQGF 1434
QY 934 LEEKQRLKIKLEGGI--KGIGKTSNNSKNLSSEPVITKAKE----GCQSDSMRQEQS- 986
DB 1435 PSRKEHKEKPELVDDLSPRRASVDSVEDSESSENPSVVRKRRTSTIGSSSSDEYKQEDSQ 1494
QY 987 -----PNANNQPEDLIQ-----GCQSD----- 1005
DB 1495 GSGEEDFIRKQIIEMSADEAGSGDEDFIRSQLKEISGVGESQKREAKGKGVAGK 1554

```

```
QY 1006 -----SSVLRMSDPSTHTNKLYPKDRVLDVVSIRSPETKCPKONSIENDIEEKVSDLAS 1059
Db 1555 HRLTRKSTSFDDAGRRHSWEDE---DEFDESPELKFRETQSQES--RELV---VA 1606
QY 1060 RGOEPTKSKTKGNDFIDDSKLASADDIGTLICKNNKPLIQEES-----DRIVSSKSA 1113
Db 1607 GGGGLRFFKTIELNSTIADKYSSSSQKKTILYFDEPELEMEESLTDSPEDSRGEGSSS 1666
QY 1114 LH-----SSVPKSTNRDRATPLSRAMDEGKLGCDSENSTLENSSDTSVTDSS 1163
Db 1667 LHASSETPGTSPTSVSLDSDSDSPSHKKGESKQOKARHRSHGPL-----LPTIEDSS 1721
QY 1164 EEDMIVONSNESTSEOFRTRE-----1184
Db 1722 EEEL-REBELKEQEKQRELEQQOKRSSKSKDKDELRAQRERRPKTPPSNLSPI 1780
QY 1185 QDVELEPLKCELVSGE-STGNCEDRLPVKGTGANGKPKSQOKLBERPVNKCSDOIKLK 1243
Db 1781 EDASPTHEELRQAAMEELHRSSCSEYSPSIESDPGEFESPEKIIIEVKVKYKLPAAVSLY 1840
QY 1244 NTDDKNNRESEKKGORTST---FQINGKDNKPKIYLKGECLKEISESRVVGNEVPK 1300
Db 1841 SPTDEQSVMQKEGVQKALKSAEBEYEMMQKPHKYKAFPAANERDEVEFEKPLYGGM--- 1897
QY 1301 VNNINKIIPENDIKSLTVKESAIRPFINGDVI-----MEDFNERNSETKSHLLSSDAE 1355
Db 1898 -----LIEDYIYESLV-EDTYNGSVGSLLTRQEBNGCFMQORGEQKVRV-----QE 1944
QY 1356 GNYRDSLETLPSTK-----ESDSTQTTTP-----SASCPESNVQOVEDMETETSEVK 1404
Db 1945 QIVDDPMOKISDLQKFEYELSHSVVPPQEDIVSSSYIIPESHEIYDGLGSMVNSTSEKK 2004
QY 1405 VT-----SSPTISE-----EESNLSNDFIDENGL-----1428
Db 2005 LLDADSAYEELMRQQOVQVTDGSPVQTTTGGDDMAESTLDFDRVQDASLTSLSGASLT 2064
QY 1429 -----PINKNNVNGESKRKTIVITEVT-----TMTSTVATESKTVIKVE 1467
Db 2065 DSTSSATLSIPDKVITQFSAEELEDEYVDTREIQDIIAHESLILTYSEPSSESATSV 2124
QY 1468 KGDKQTVVSTENCACKSTVTTTTVTYTKLSTPSTGGSVDLISVKEQSKTVVTVTVDSLT 1527
Db 2125 PSDTSLSTISSYCTDSSPVTLLDSTLVYT-EPADVMTKFKDEEISSTYFFGSII 2183
QY 1528 TTGCTLVISM--TVSKEYSTRDKVKLMKFS-----RPKTRSGTALPSYR 1570
Db 2184 DYPEDISVSLDRTIMPESRTNEDRIVLSFGMAPSVSVESVGTKPERQADTISTDLPISE 2243
QY 1571 KFTVTKSKSIFVLPNDLKLARKGGIREVPYFNYNNAKPAOLDIWPYSPRPTFGITWRY 1630
Db 2244 KDLIKG-----RKETGDGIILEV-----LDAYK-----DK 2268
QY 1631 RLQTVKSLAGVSLMLRLWLASLWDDMAKVPPGGSGTRETSETTEITTEIKRRDVG 1690
Db 2269 REESEAEELTKISL-----PEPGLAQAPSS-----VTAQIKEQHVSP 2305
QY 1691 YGIRFEYCIRKIICPIGVPETP-----KETPTPQKGLRSSAL-----RPKRPE 1735
Db 2306 HVSQK-----ISQEKPTYRLPSGLSPVSTHPSKSRPFRSSSLDISAQPPPPPP 2356
QY 1736 PKQTGPVLIETWVAEELELWEIRAFARVEKEKAQAQAVEQAQKRLKQKPTVIATSTTS 1795
Db 2357 PPPPSP-----STSSPPPTPLPPPAT 2378
QY 1796 PTSSTSTISPAQKVMVAPISGVTGTGKMWLTTKVGSPTATVTFQKNKFHOTFATWV-- 1853
Db 2379 PKPPT-----YPKKLAVA---ATVISTT--IVTHVDALTWVEAAARRSNGLPATKCA 2429
QY 1854 -----KQGSNSGVV-----QVOQKVLGIIPSSSTGTSTQO- 1882
Db 2430 IAPPPVPKPSQITGLVFTHRPEAIKPPPIAPKPAVPIPVTTQKPTDTCFKPTGLSLTS 2489
```

RESULT 13
BA2B_HUMAN
ID BA2B_HUMAN STANDARD; PRT; 1972 AA.
AC Q9UIF8; Q96EAL; Q96SQ8; Q9P252; Q9Y4N8;

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Bromodomain adjacent to zinc finger domain 2B (hwalp4).
 GN BAZ2B OR KIAA1476.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC
 RC TISSUE-Testis;
 RX MEDLINE=20130112; PubMed=10662543;
 RA Jones M.H., Hamana N., Nezu J., Shimane M.;
 RT "A novel family of bromodomain genes.";
 RL Genomics 63:40-45(2000).
 RN [2]
 RP SEQUENCE OF 1-776 FROM N.A. (ISOFORM 1).
 RC
 RC TISSUE-Melanoma;
 RA Ansoerge W., Wirkner U., Mewes H.-W., Weil B., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-799 FROM N.A. (ISOFORM 2).
 RC
 RC TISSUE-Skeletal muscle;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ussid T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Maira M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP SEQUENCE OF 427-1215 FROM N.A. (ISOFORM 3).
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
 RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
 RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K.,
 RA Masuho Y., Kanehori K.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 753-1972 FROM N.A. (ISOFORM 1/2/3).
 RX MEDLINE=2027482; PubMed=10819331;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:143-150(2000).
 RN [6]
 RP SEQUENCE OF 1524-1972 FROM N.A. (ISOFORM 1/2/3).
 RC
 RC TISSUE-Testis;
 RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: May play a role in transcriptional regulation
 CC interacting with ISWI.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing: Named isoforms-3;
 CC Comment-Experimental confirmation may be lacking for some
 CC isoforms;
 CC

CC Name=1;
 CC IsoId=Q9UIF8-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9UIF8-2; Sequence=VSP_000553;
 CC Note=Inferred from Ref.2;
 CC Name=3;
 CC IsoId=Q9UIF8-3; Sequence=VSP_000554;
 CC Note=Inferred from Ref.3;
 CC -!- TISSUE SPECIFICITY: Expressed at varying levels in several
 CC tissues, whereas a smaller transcript was expressed specifically
 CC in testis
 CC -!- SIMILARITY: BELONGS TO THE WAL FAMILY.
 CC -!- SIMILARITY: Contains 1 bromodomain.
 CC -!- SIMILARITY: Contains 1 DDT domain.
 CC -!- SIMILARITY: Contains 1 methyl-binding (MBD) domain.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
 CC frameshift in position 731.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB032255; BAA89212.1; -;
 CC EMBL; AL834381; CAD39044.1; ALT_INIT.
 CC EMBL; BC014576; AAH12576.1; ALT_FRAME.
 CC EMBL; AK027612; BAB55231.1; ALT_INIT.
 CC EMBL; AB040909; BAA96000.1; -;
 CC EMBL; AL080173; CAB45759.1; -;
 CC PIR; T12495; T12495.
 CC HSSP; Q2831; LB91.
 CC Genew; HGNC:963; BAZ2B.
 CC MIM; 605683; -;
 CC InterPro; IPR001487; Bromodomain.
 CC InterPro; IPR004022; DDT_dom.
 CC InterPro; IPR001739; Methyl-CpG_bind.
 CC InterPro; IPR001965; Znf_PHD.
 CC Pfam; PF00439; bromodomain; 1.
 CC Pfam; PF02791; DDT; 1.
 CC Pfam; PF01429; MBD; 1.
 CC Pfam; PF00628; PHD; 1.
 CC PRINTS; PR00503; BROMODOMAIN.
 CC SMART; SM00297; BROMO; 1.
 CC SMART; SM00571; DDT; 1.
 CC SMART; SM00391; MBD; 1.
 CC SMART; SM00249; PHD; 1.
 CC PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 CC PROSITE; PS50014; BROMODOMAIN_2; 1.
 CC PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
 CC PROSITE; PS50016; ZF_PHD_2; 1.
 CC KW Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;
 CC Nuclear protein; DNA-binding; Alternative splicing.
 FT DOMAIN 549 601 MBD.
 FT DOMAIN 891 953 DDT.
 FT ZN_FING 1735 1785 PHD-TYPE.
 FT DOMAIN 1881 1951 BROMODOMAIN.
 FT DOMAIN 45 69 SER-RICH.
 FT DOMAIN 73 79 POLY-GLU.
 FT DOMAIN 399 470 ASP-GLU-RICH.
 FT DOMAIN 641 672 ARG-RICH.
 FT DOMAIN 687 865 COILED COIL (POTENTIAL).
 FT DOMAIN 706 875 LYS-RICH.
 FT DOMAIN 1100 1143 ASP-RICH.
 FT DOMAIN 1138 1179 COILED COIL (POTENTIAL).
 FT VARSPPLIC 437 534 Missing (in isoform 2).
 FT VARSPPLIC 593 626 /FTId=VSP_000553.
 FT VARSPPLIC Missing (in isoform 3).
 FT CONFLICT 137 137 S -> F (IN REF. 1).

FT	DOMAIN	1968	1979	POLY-SER.								
FT	VARSPLIC	2333	2371	Missing (in isoform Short).								
FT				/FTIQ-VSP_003411.								
FT	CONFLICT	1952	1952	I -> T (IN REF. 2).								
FT	CONFLICT	2090	2090	A -> P (IN REF. 2).								
SQ	SEQUENCE	2453	AA: 270640 MW: 52208B40382F7E6A	CRC64:								
Query Match												
Best Local Similarity 18.9%; Pred.No.1.9e-06;												
Matches 434; Conservative 302; Mismatches 782; Indels 782; Gaps 108;												
QY	298	HKVPGVDCVAEIQKKNPYIRH	----	EPICYDRSRKRYWFLNRLRIIEED	----- 343							
DB	378	HEISEIIDGLSEOENNEKQMRQLSVIPPMFMDAERQVVFNNGLMEDPMKYVKDQRFM	437	----	----							
QY	344	----	TENENE	-----	KKIYVSTKVQLAELIDCLDKDYAEALCKILEEMREIHR	390						
DB	438	NVMTDHEKEIFDKFIQHPRNFGFLIASIYERKSVPDVCVLYY	----	LTKKNENYKALVRR	493	----						
QY	391	HMDITDLTKAGSNKSFLLAAANEETLESIRAKGIDNVKSPETEEDKDNETENDSKD	450	----	----							
DB	494	NY	-----	CKRRGNQOIARPSQOEKVEKEEDKA	-----	EKTEKEEEKKDDERK	538					
QY	451	AENREEFQSLKSDDKTDDPQKQSEYGVDFKSEKSNSELSPEGAGKAGSGSTR	510	----	----							
DB	539	DDK	-----	EDSKETTEKDRTEATA	-----	EEPEE	-----	REQVTPGRKTAN	-----	SQGRGKR	-----	584
QY	511	IITRLNPDSKLSQKSOVAAAAHEANKLFKEGKEVLYVNSQGEISRLSTKK	EVIMKG	569	----	----						
DB	585	VTR	-----	SMTSEAAAAANAATAEPPPPPPPISTEPVTSRWTSEEMEYAKG	638	----	----					
QY	570	NINNYEKLGOEGKYRVYHNOYSTNSFALNKHQREDHDKRHLAHKFLCLTPAGEFKWGS	629	----	----							
DB	639	LVEHGRNWAATAKMGVCTKSEAQCKNFYNY	-----	KRRH	-----	-----	672					
QY	630	VHGSKVLITSLRLTITTOLENNIPSPFLHPNASHRANWIKAVQMCKSPREFALALALE	689	----	----							
DB	673	-----	NLDNLLQO	-----	HKQKASR	-----	-----	KPRE	-----	ERDVSQ	697	
QY	690	CAVPVVMPLPIWREFLGHTLRHMTSTIEREEKVKKKKEQBEETMOQATWVKYTFPV	749	----	----							
DB	698	C	-----	ESVAST	-----	VSAQEDEDIASNEENPEPDESGAENSSDTE	SAPS	739	----	----		
QY	750	KHQMVKQKGBE	----	YRVTVGGVSWISKTHV	-----	YRFVPLPGNTNVNRYKSL	796	----	----			
DB	740	PSPVEAAKSSDSENAAASRGNTPEFALEATTDPAPCASPSSAVP	----	TTKPAERESV	795	----	----					
QY	797	EG	-----	TKNMENDESCKRCS	----	RSPKKIKIEPDSKEKVGSDAAKG	ADQ	844	----	----		
DB	796	EAQVTDASAETAEPMDVDHEECGAEGSSVLDPPAPTKADSVDPDMQVPENTASKGEGDA	855	----	----							
QY	845	NEMDI	----	SKITEKKDQDV	-----	KELLDSDSKPKCEE	-----	PMEV	879	----	----	
DB	856	KERDLESTSEKTEARDEVDVVAEOIERPEPOSDDSSATCSADSGVDGEPEQRQVFMDA	915	----	----							
QY	880	DDMKTE	-----	SH	-----	VNCOESSQVDVVNVSEGFHLRTS	911	----	----			
DB	916	KPSLLTPGSLISSPIKPNLLDLPQLQHRAAVIPMVSCPCN	IPIGTPVSGYAL	971	----	----						
QY	912	YKKTSSKLDGLLERIKQFTLEEKQRLKLEGGIKGIGKTSNSSKN	-----	LSES	966	----	----					
DB	972	YORHIKAMESALLE	-----	EQRQREQVDLECR	-----	SSTSPCSTSKSPNREWEVLQPA	1022	----	----			
QY	967	P	-----	VITKAKGQSDSMREQSP	-----	NANNQPEDLIQCGSQSDSVLRMSDPS	1015	----	----			
DB	1023	PHQVITNPEGVRLPTTRPPPLIPSSKTTVASEKPSFIMGSIQSGTGGYLS	-----	S	1080	----	----					
QY	1016	HTTNKLPKORVLDDVSRISPEYKCPKQNSIENDIEKVSDIAGSQBPPTKSKTKGNDFF	1075	----	----							
DB	1081	H	-----	NQAYPQ	-----	EAPKPSVSGISLGL	-----	PROQUESTRA	-----	-----	1110	
QY	1076	IDDSKLASADDIGTLICKNKXPLI	-----	QESDITVSSSKSALHSSVPKSTNDROATPLSR	1132	----	----					

Db	1111	-----	APLVTKQEEFS	-----	PRSQNSQPEGLLVR	1136				
Qy	1133	AMD	-----	FEGKLGCDSESNSTLENSDVTVS	IODSSEDMIVONSNESISEQFRTREDDVEV	1189				
Db	1137	AOHEGVVRGTAGAVQEGSITRGTPASKIS	-----	-----	VET	1168				
Qy	1190	LEPLKCELVSGSTGNCEDRLPVKG	----	TEANGKPPQOQKLEERPVPVKCSDQIKLKNTT	-----	1246				
Db	1169	ISSLRGSIQGTGA	-----	LPQAGIPTAALYKGPVSRMP	IESSPEKVREEAASKGHVI	1222				
Qy	1247	DKNN	-----	ENRESEKKGORT	-----	STFOINGKDNPKI	-----	YLKGECLKEISERSVVG	1295	
Db	1223	YEGKSGHILSYDNINKAREGTRSPRTAHMSLUKRYEAVEGSIKOGMSRESVPVAPLEG	1282	----	----	----	----	----	----	
Qy	1296	NVEPKVNNINKIIPENDIKSLTVKESAIRPFFINGVIM	-----	EDFNE	-----	-----	-----	R	1339	
Db	1283	LI	-----	CRALPRGSPHSDLKERTV	-----	LSGSIMOGTPRATAESFEDGLKYPQIKR	1331	----	----	
Qy	1340	NSETKSHLLSSDAEGNRYRDSLETLPSTKESDSTQTTPTSASCSPESNSVNOVEDMEIET	1399	----	----	----	----	----	----	
Db	1332	ESPPIRA	-----	FEGAITKGKPYDGITTI	-----	KEM	-----	GRSIIHPIPRODILIT	1372	
Qy	1400	SEVKV	-----	TSSPITSEESNLNDFIDENGLPINKNENYNGSRRKRTVITEVTTMTS	1454	----	----	----	----	
Db	1373	QESRKTPVWQSTRPII	-----	EGSIS	-----	OGTPI	-----	KFDNNSGQSAIKHNKSLITGPS	1422	
Qy	1455	TVATESKTVIK	-----	VEKDKQ	-----	TVVSTENCAKSTV	-----	-----	1486	
Db	1423	KLPRGMLEIVPENIKVVERGKYEDVKAPEVRARHTSVSSGSPVLRLTHEAPKAQLSP	1482	----	----	----	----	----	----	
Qy	1487	-----	TTTTTTVTVKLSTPSTGGSV	-----	DIISVKEQS	-----	KTVVTVTVTDSLTTTG	-----	1530	
Db	1483	GLYDDSSARTPVSYQNTISRGSPMNRISDVSSSKASHERKSTLTPQRSIPAKSPV	1542	----	----	----	----	----	----	
Qy	1531	-----	GTIVT	-----	SMTVSKEYSTRDKVKL	-----	MKFSRPKKTSGTAL	-----	1566	
Db	1543	PGVDPVIVSHSPDPHRRSSAAGEVYRSHLPTHLDPAMPFHRALDPAAVLLQRLSPTPG	1602	----	----	----	----	----	----	
Qy	1567	PS	-----	YKRFVTKSTKISFVLPNDLLKLARKGIREVYFNNAKP	-----	ALDIWVPSPR	1621	----	----	
Db	1603	YPSOYQLYAMNTRQITL	-----	NDYITTSQOMQVNL	-----	PDVTRGLSPREPQGLG	PPAPVR	1656	----	
Qy	1622	PTFGITWRYRLQTVKSLAGVSLMLRLMLASRLWDDMAAKVPP	-----	GGGSTRTETSE	1674	----	----	----	----	
Db	1657	GIIDL	-----	-----	-----	NMPPTILVPHAGGTSTPPMDRI	1684	----	----	
Qy	1675	TEITTEIIRKRDVGPYIRFEYCIRKIICPIGVPETPKETPTPQKGLRSSALRKPE	1734	----	----	----	----	----	----	
Db	1685	TYIPGTQVT	-----	-----	-----	FPPRP	-----	YNAASLSPGHP	1709	
Qy	1735	TPKQGTGVIETWVAEELELWEIRAPAEVEKEKAQAVEQAKKRLQOQKPTVIATSTT	1794	----	----	----	----	----	----	
Db	1710	-----	THLA	-----	AAASAER	-----	EREREREKERERERER	-----	1740	
Qy	1795	SPTSSTTSPACKVMVAPISGVTGTWKVLTTKVGSPATVTFOONKNFHOTFATWK	1854	----	----	----	----	----	----	
Db	1741	-----	ERERIAAPADLYLRPGE	-----	QFGRPGSHGV	-----	-----	-----	1770	
Qy	1855	QGSNSGVVQVQKVLGIIPS	-----	STGTSCQQTFTSQPRTATVTPRPNTSGSGGTTNS	1910	----	----	----	----	
Db	1771	--RSPSPSVRTQETILQORPSVFOGTNGTS	-----	VITPLDP	-----	TAOLRIMPLPSG	-----	-----	1817	
Qy	1911	QVITGPOIRGMVIRTPLOOSTLGLKAILRTPVNVQGPQOVMTQIIRGPQVSTAVSAP	1970	----	----	----	----	----	----	
Db	1818	-----	GPISISQGLPASR	-----	YNTAADALAL	-----	-----	-----	VDAAASAP	1848
Qy	1971	NTVSSTPGQSLTSATSTNTQSSASQPPRPOQSQVKLTMAQLTQLTQGHGNOGLT	2030	----	----	----	----	----	----	
Db	1849	QMDYSKTESKESHEARLEENLRSSAAVSEQQLEQKNLEVE	-----	-----	-----	KRSVQVCV	1898	----	----	
Qy	2031	QGOQTTGQIQIOLIPQGVTVLPGPQQQLMQAAMPNGTVQORFLPTPLATTATTATTTTTS	2090	----	----	----	----	----	----	
Db	1899	TSSALPQSKAQ	-----	PHASVTVSEAGKD	-----	KGPPKRSRYEELTRGKTTIITANFIDVIT	1954	----	----	

THIS PAGE BLANK (USPTO)

```

Db      1531 S-----NSDPHD-----IPESKDCVLT-----SEEMFSKDK 1557
QY      1889 PTAIVTIRPNTSGSGTTSNQVITGPQIRPCMTVIRTPLOOSTLKGKAIIRTPVMVQPG 1948
Db      1558 TFIVRQSHDEISVSMDSASQMLNEEQLEDMDRQELVRYQOEHOQATELLRLQAHMRQME 1617
QY      1949 APQQVMTQIIRQPVSTAVSAPNTVSSYTPGOKSLTSATSTSIQSSASQPPRPOQGVKL 2008
Db      1618 ROREDOEQL---QEIKRLNRLQARSSIDNENLYSERERVILLEALKQLSLAGREKL 1674
QY      2009 TMAQLTQTLQGHGNGOGLTVIQQGQTTG-----LQLIPOGV--TVLPGGQOLMOA 2060
Db      1675 CCELRNSSLTQNGN-----ENQGEVEEOTFEKELDRKPEDVPPEILSNERYALQKA 1727
QY      2061 AMPNGTVQRFLEPTPLATTATTAST-----TTTTVSTTAAGTGEOROSKLSQPMOVH 2111
Db      1728 ---NNRLKILLEVYKTTAAVEETIGRHVLGILDRSSKSSQSSASLIWRSEASVSKCVH 1784
QY      2112 QDKTLP-----PAQSSSVGP-----AKAQPQTAPQARPQPTQPOSPAQE----- 2153
Db      1785 EEHRTVDESIFSYSGSDMPRNDINWMSKVTEEGTELSQRLVRSFGAGTEIDPENEELML 1844
QY      2154 -VOTQEVQTTQVSHVPSEAOPTHAOSSKPOVAAS-----QPQSNVQOSPVRVQSP 2207
Db      1845 NISSRLQAAVEKILLEAISSETSSQLEHAKVQTQELMRSPROKQEATESLKCOEELRERLH 1904
QY      2208 SOTRIRPSTPSQLSPQ-----OSQVQTTT-----SOPPIQPHTSLOIPS 2248
Db      1905 EESRAREQJAVELSKAEGVIDGYADEKTLFEQIOEKTIDIRLEQELLCASNRLQOELEA 1964
QY      2249 QGQPSQPOVQOSTOTLSSGQTLNOVSSPSRPQIQPOQPOVIAVPQLOQOVQVLSQ 2308
Db      1965 E-----QQQIQEERELLSR---QKEAMKAEAGP---VEQ-----QLLQETELMK 2003
QY      2309 IQSQVVAQIAQOQSGVPQOIK-LQLPIQIQSSAVQTHOIQNV-----VTVQAASVQEQLO 2363
Db      2004 EKLEVQCAEKVRDLDQKQKALEIDVEQVSRLFEOEKNTLMELMDLRQOQNALEKQLE 2063
QY      2364 RVQOLRQOQKKQOQIEIKREHTLQASNQSEIOLK-QVVMKHNA-----VIEHLKOKK- 2416
Db      2064 KHRKFLDEQAIDREHEDVFOEIQKLEQOLKVVFPFQPISEHQTREVEQLANHLKEKTD 2123
QY      2417 --SMTPAEREENQRMIVCNQMYILDKID---KEEQA--AKKKREESVEOK----- 2463
Db      2124 KCSLELLSKEQLQORDI---QERNEETEKLFRVRELEQALLVEDRKHFGAVEAKPELSLE 2180
QY      2464 -----RSQNATKLSALLFKHKEQLRAELILKRALLDK-DLQIEVQ----- 2503
Db      2181 VOLQARDAIDRKEKEITNLEQL-----EofRELENKNEEVQOQLHMQLEIOKKESTRL 2236
QY      2504 EELKRLKIKKEKDLMLAQATAVAAPCPVTPVLPAPPAPPPPPPGVQHTGLLSTP 2563
Db      2237 OELEQENKLFKD-DMEKLGIA-----IKESDAMSTQ 2266
QY      2564 TLPVASQK-----RKREEDK-----SSSKSKKKMISTTSKETKDKTKLYCICKTPYDES 2614
Db      2267 DQHVLFGEKFAQIIQOEKEVEIDQNEQVTKLOOQKITTDNKVIEEKNELIRDET----- 2321
QY      2615 KFYIGCDRCNWNHGRVGLQSEAEILIDEYVCPQOSTEDAMTVLTPTEKDYGLKRV 2674
Db      2322 --QIEC-----LMSDQECVK-----RNREEIEQNEVIEKLEQOQELANI 2358
QY      2675 --LRSLOAHKMAWPFLEPVPNDADPYGYVKEKPMDLATMEERVQRRYVEKLETFVADMT 2732
Db      2359 GQKTSMAHSL-----EADS-----LKHQLDVIAEKLALQEQVETANEFTFMK 2405
QY      2733 KIFDNCRYNPSDPFYQCAEVLESFFVOKLKGFKASRSHNNKLOS 2778
Db      2406 NVLKETNF-----KMNQLTQELFSLKRERESVEKIQS 2437

```

```
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171.311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 3899
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-4

Query Match      2.7%   Score 384.5; DB 15; Length 3899;
Best Local Similarity 17.2%; Pred. No. 1e-10;
Matches 483; Conservative 475; Mismatches 1035; Indels 813; Gaps 115;

QY    305 DCVAEI-----QKNKP--YIRHEPIGYDSRRKYW-----FLNRLIIE 341
Db    113 :|::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~
QY    342 EDTENENNEKKIIWYSTKVQLAELIDCL-----DKDYWEAECLCKIL 381
Db    173 ELNRELEEMRVTYGTEGLQLOQEFAALKQRDGIITOLTANLQARREKDETMREFLELT 232
QY    382 EEMRE-BIH-RHMDITEDLNKARGSNKSFLAAANEFI-----LESIRAKKG 426
Db    233 EQSKLOIQFQOLQASSETLRNSTSHSTAAADLLQAQQILTHQQOOLEQDHLEDYQKKKE 292
QY    427 DI-----DNVKSPEETEDKNETENDSKDAEKNRE-----EPEDOSLEKDS 467
Db    293 DFTMQISFLQEKIK-VYEMODK-KVENSKEEIQEKETIIEELNTKIEBEKTLEUKD 350
QY    468 DKTPD-----DDPQGKSGVDGFKSEKNGELSESPGAG--KGASGSTRIITRLRNP 518
Db    351 KLTTADKLGLLQEQIVQKNOEIKNMKLELTNSQKERQSSEEIKQLMGTVEELOKRNHK 410
QY    519 DSKI-----SOLKSQOVAAAHAEA-NKLFEKGEKVLVYNVSQGEISRSLTKKEVI 566
Db    411 DSOFETDIVORMEOETORKLEQLRAELEDEMYGOQIVQMQLIRHQMAQMEEMKTRHKGE 470
QY    567 MGNINNYFKLGOGBKRYVYHNOYSTNSFALNK-HOHREDHDRRHHLAHFCLTPTAGEF 624
Db    471 MENAIRSYNI-----TVNQDQKLMNVAINELNIKLDQTNQOKERLUKEEL-----516
QY    625 KWGSVHGSKVLATISTLRLLITQLENNIPSSFLHPNWASHRANMIKAVOMCSKPREFALA 684
Db    517 --GLILEECALQROLEDLVEEL-----SFSREQ-IQAROTIAOE-SKLINEAKHS 564
QY    685 LAILECAVKPVMLPIWREFLGHTRLHRMTS-----IEREEKYKKKKKEKQOEEETMOQA 740
Db    565 LSTVEDLKAEITVASSESKELELKHAEVTNYKIKLEMLEKEKNAVLDRMAESQAEALER 624
QY    741 TWKYTFPPVKHOWWKQGE---ERYVTVGCGWSISKTHYVFVPKPLPGNTNVNRYKLSLE 797
Db    625 LRTQLLFSSHEELSKLEDLEIEHRIN-----TEKLDNLGIHYKQOID 668
QY    798 GTKNNDENMD--ESDKRKCSRSPKTIKIBPDSEKDEVKGSDAAKAGD-----QNEM 847
Db    669 GLQNMESQKITETWQFEKDNLITKONQLILEISKDLQOOSLVNYSKSEEMTLQINELQKEI 728
QY    848 DISKITEKK----DQDYKEL-----LSDSDRPCKPEPMEV 879
```

QY	1923	TVIRTPLOQSTLGLKAIIRTPVMVQPCAOQVMTQIIRQPVSTAVSAPNTVSSTPGQKSL	1982
Db	3909	TTPTPTGTGTPTTTPIITTTTTP-TPTPTGTQ-----TPTTPIITTTTTPPTPTPTGT	3963
QY	1983	TSATSTNSIQSASQP-PRPOGOVKLMAQLTOLTQGHGNGOGLTVVIOGOGQTGQLQ	2041
Db	3964	QTPPTTPIITTTTTPPTPTGTGTPTTTPIITTTT-----VTPPTPTGTGTPTTT	4015
QY	2042	LIPOGVTVLPFGQOQLMQAAMPNGVQRFLEPLATT-----ATTASTTT	2086
Db	4016	PIITTTTTPPT-----TPTGT-QPTPTTPIITTTTTPPTPTGTGTPTTTPIIT	4066
QY	2087	TVVSTTAAGTGQRQSKLSPQMVHODKTLPPAQSSVGPAPAKAQPTQAPSARPQPTQP	2146
Db	4067	TVVTPPTPTGTGTPT-----TPTTTTTPPTPTPTGTGTPTTTPIITTTT	4114
QY	2147	QSPAQEVOTQPEVOTQTVSSHPVSEAOPTHAQSKKPQVAAQSQPSQSNVQSGSPVRVQS	2206
Db	4115	VTPPTPTGTGTPTTTPIITTTTTPPTPTPTGTGT-----TPTTPIITTTTTPPTPT	4172
QY	2207	PSOTRIRPS---TPSOLSPGOQSOVOTTTTQPIQPHSTSLQIPSGOQPQ-----SQPOV	2258
Db	4173	PTTPTTTTTPPTPTGTGTPTPTPTGTGTPTPTPTPTPTPTPTPTPTPTPTPTPTPT	4232
QY	2259	QSSTQTLSSGQTLNQVSVSSPSRP	2282
Db	4233	ESTTLLSTLPPAIENTSTAPPSTP	4256

RESULT 14

US-10-025-380-1068

; Sequence 1068, Application US/10025380

; Publication No. US20020182191A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Lodes, Michael J.

; APPLICANT: Secrist, Heather

; APPLICANT: Benson, Darin R.

; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Stoik, John A.

; APPLICANT: Wang, Tongtong

; APPLICANT: Jiang, Yugu

; APPLICANT: Smith, Carole L.

; APPLICANT: King, Gordon E.

; APPLICANT: Wang, Aijun

; APPLICANT: Clapper, Jonathan D.

; APPLICANT: Skeiky, Yasir A. W.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Carter, Darrick

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE

; FILE REFERENCE: 210121.47IC14

; CURRENT APPLICATION NUMBER: US/10/025.380

; CURRENT FILING DATE: 2001-12-19

; NUMBER OF SEQ ID NOS: 1129

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1068

; LENGTH: 5179

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-025-380-1068

QY	1236	PSKESDSTQTTTSPASCPESNVQVEDMEIETSEVKKVTSSPITSEESNLSNDFIDE	1425
Db	3460	PTTPTTTTTPPT	3508
QY	1426	NGLPINKENV-----NGESKRKVTITEVTTMTST---VATESKTVIKVEKGDKQTVV	1475

RESULT 15

US-10-171-311-4

; Sequence 4, Application US/10171311

; Publication No. US2003008720A1

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Chen, Yan

; APPLICANT: Zhao, Xumei

; APPLICANT: Monahan, John

```

; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-922-217-1068

```

Query Match	2.7%	Score 391;	DB 9;	Length 5179;
Best Local Similarity	22.5%;	Pred. No. 6.8e-11;		
Matches 221; Conservative	91;	Mismatches 418;	Indels 254;	Gaps 33;
QY	1366	PSTKESDSTOTTPSASCPSNESVNOVEDMEIETSEVKVKTSSPITSEESNLNDPIDE	1425	
DB	3460	PTTPTPTTTTVPPT---PPTGTGTPTPTPTTTTVPPTPTGTGTPTPT	3508	
QY	1426	NGLPINKNENV-----NGESKRKTIVITEVTTMTST---VATESKTVIKVERGDKOTVV	1475	
DB	3509	--TPTITTTTVPPTPTGTGTPTPTPTTTTVPPTPTGTGTPTPTPTTTTVPPT	3566	
QY	1476	SSTENCASKSVTTTTTIVKLSPTSGGSDIILSVKEQSKTVVTTTVDLSLTWGGT--L	1533	
DB	3567	PPTGTGTPTPTPTTTTVPPTPTGTGTPTPTPTTTTVPPTGTGTGTPTGTPTPTTI	3626	
QY	1534	VTSMVTSEKYSTRDVKVLMKFSRPKPTRSGTALPSYKRFVTKSTKKSIFVLPNDDLKLA	1593	
DB	3627	TTTTTVP-----PPTPTGTGTPTPTPTTTT	3654	
QY	1594	RKGIREVPYFNNAKPAIDIWYPPRPRTFGITWRYRLQTVKSLAGVSLMLRLWASLR	1653	
DB	3655	-----VTPPTPTGT-----QTPTTPTIT	3673	
QY	1654	WDDMAAKVPPGGGSTRTET--SETEITTEILKRRDVGVPYGRICEYCKRIICPGVP--E	1710	
DB	3674	---TTTIVTPTPTGTGTPTPTPTTTTVPPTPT--PTGTQ-----TPTTPTITT	3720	
QY	1711	TPKETPTPQKGLRSALRP---KRPEPKGTG-----PVLIETWVAEELELWEI	1758	
DB	3721	TTTIVTPTPTGTGTPTPTTTTVPPTPTPTGTGTGTPTPTTTTPTT	3771	
QY	1759	RAFAERVEKAQVSEQAKRLEBQOKPTVIASTSPS---STWSTISP-----AQK	1809	
DB	3772	-----PPTPTGTGTPTPTTTTVPPTPTPTTTTVPPTPTGTGT	3804	
QY	1810	VMVAPISGSVT-----TGCTKMVLTKVGSPATVTFQOKNKHOFATVWKQGSNSGV	1862	
DB	3805	PTTPTPTTTTVPPTPTGTGTPTPTTTTVPPTPTPTPTPTPTPTPTPTPTPTPT	3861	
QY	1863	VQVQKVLGIPSTGTSQOTFTSFOPRTATVIRNPNTSGSGGTTNSQVITGPQIRPGM	1922	
DB	3862	-----VTPPTPTGTGTPT--TPTTTTVPPTPTPTGTGTPTTPTIT---TPTTV	3908	
QY	1923	TVIRTLQOSTLGKALIRPVMVQGAQOVMTQINGQVSTAVSNAPNTVSSTPGQKSL	1982	
DB	3909	TPPTPTGTGTPTPTTPTTTTVP--TPTPTGTQ---TPTTPTTTTVPPTPTPTGT	3963	
QY	1983	TSATSTSNIOSSASQP--PREQOGVKLTMQAQLTQGHGNGOGLTVVIOGOGTQQLQ	2041	
DB	3964	QPTPTPTPTTTTVPPTPTPTGTGTPTPTPTTTT-----VTPPTPTGTPTPTT	4015	
QY	2042	LIPQGVTVLPGPQQQLMQAAMPNGTVQRFLETPLAT	2086	
DB	4016	PTTTTTPVTP-----TPTGT--QTPTPTPTTTTVPPTPTPTGTPTPTTPTT	4066	
QY	2087	TVVSTTAAGTGEORQKSLSPQMVOHQDKTLPPAQSSSVGPAKAPQTAQSAARPQOTQ	2146	
DB	4067	TTVTPPTPTGTGTPT-----TPTTTTVPPTPTPTGTGTPTPTTPTT	4114	
QY	2147	QSPAQEVOTOPEVOTGTTVSSHPSEAOPTHAQSKSPQVAAQSQPSNVQGSPPVRVQS	2206	
DB	4115	VTPPTPTGTQNPPTPTPTTTTVPPTPTGTGT--PTTPTTTTVPPTPTPTGTGT	4172	
QY	2207	PSQTRIRPS---TPSOLSPGQQSQVQTTTSSQPIQPHSTLSQIPSGQPO-----SQPQV	2258	

[illegible]

Query Match	2.7%	Score 391;	DB 10;	Length 5179;
Best Local Similarity	22.5%	Pred. No. 6.8e-11;		
Matches	221;	Conservative	91;	Mismatches 418; Indels 254; Gaps 33;
Qy	1366	PSIKESDSTOTTPSASCPSESNVQVEDMEIETSEVKKVTSSPITSEESNLSDNFIDE	1425	
Db	3460	PTTTPITTTTVPPT---PPPTGQTPTTPIITTTTVPPTPTGQTPTT-----	3508	
Qy	1426	NGLPINKNNV-----NGESKRKVTITEVTMTST--VATESKTVIKVEGDKQTVV	1475	
Db	3509	--TPITTTTVPPTPTGQTPTTPIITTTTVPPTPTGQTPTTPTTPTTPTTPTT	3566	
Qy	1476	SSYENCAKSNVTTTIVTKLSPSGGSDVILISVKEOSKTVTTVTTDSLTITGCT--L	1533	
Db	3567	PTPTGQTPTTPTTPTTPTTPTPTGQTPTTPTTPTTPTTPTPTGQTPTTPTTPI	3626	
Qy	1534	VTSMTVSKEYSTRDKVKLMKFSRPKPTRSGTALPSYRKFTKTKKSIFVLPNDDLKLA	1593	
Db	3627	TTTTTVI-----PTPTGQTPTTPTTPTT-----	3654	
Qy	1594	RKGIREVPYFNNAKPALDINWPSRPPFGITWRYRLOTVKSLAGVSLMLRLWASUR	1653	
Db	3655	-----VTPTPTPTGT-----QTPTTPTPT-----	3673	
Qy	1654	WDDMAAKVPPGGSTRDET--SEETITTEILIKRRDVGPGVIREYCIKRIICPGVP--E	1710	
Db	3674	--TTTTVTPPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT	3720	
Qy	1711	TPKETPTPQRKLRSSALRP-----KRPTPKQTG-----PVIIETWVAEELELWEI	1758	
Db	3721	TTTWTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT	3771	
Qy	1759	RAFAERVEKEAQAQVEQAKKRLBQOKPTVIATSTTSPTS--STTSTTSP-----AQK	1809	
Db	3772	-----PPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT	3804	
Qy	1810	VMVAPISGSVT-----TGTKMVLTKVGSPATVTFQOKNPHQTFAFWVKOGQNSGV	1862	
Db	3805	PTTPTTPTTPTTPTTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT	3861	
Qy	1863	VQVOQKVLGIIPSGTSTSQOTFTSFORTATVIRPNTSGSGGTTNSQVITGQIRPGM	1922	
Db	3862	-----VTPPTPTGQTPTT--PPIITTTTVPPTPTGQTPTTPTTPTTPTTPTTPTT	3908	

Db	787	-----IMTSTASTSKTSTVSLSDSVASAKSLSTSESNVSSTSTSLVN-----	831
Qy	911	SYKKTKSKSLDGLLER--RIKQFTLEKQREKLEKIGGKIGTKTSTNKKSLNESPV	968
Db	832	---SQVSSSMGSGVSKTSLSDF--ISNSSTSEKSE-----SVSTSTSDSLRTSTS	878
Qy	969	ITKAGCGOSDMRQEQPNANNDOPEDLIOCCSOSDSVLRMSDPTHTTNKLYPKDRVL	1028
Db	879	LSDSVSMSTSGSLSKQSLSLSTSDASTSQSVSDSTNSI-----STSESLSSEGSTS	932
Qy	1029	DDVSIRSPETKCPKQNSIENDIEEKVSDLAGRQEPKTKYKNGDNFFIDDSKIAS	1083
Db	933	ESISI-----SNSISNSVASTSKLESQSTISLSLTS-----DSKSMSTSESL	975
Qy	1084	ADDIGTLICKNKPLIOBEDPIIVSSKSALHSSVPKSTNDRDATP--LSRAMDFEGLG	1141
Db	976	SDTST-----SDSVSGSLSVASQSVSTSDSMSTSEMI SDSMSTSGSLA	1022
Qy	1142	CD-----SESNSTLENSDTSVQDSSEEDMIYONGNESISEOFRFREODVE	1188
Db	1023	ASDSKMSVSSMSTSGSTSESLSDSISLSDSDSKLSLSTSQSGSTSTSTSTSSVR	1082
Qy	1189	VLEPLKCELVSGESTCNCEDRLPVKGTGANGKKPQOQKKLEERPVKNCSDQIKLNTDK	1248
Db	1083	MSE---SQSTSGSMSTSQSDSTSI-----TSFSD-----STSDS	1114
Qy	1249	KNNENRESKKGQRTSTFOINGKDNKPKIYKGECLKIEISBRVVGNEVPKVNINKII	1308
Db	1115	KSASTASSEISQSVST-----STSGSVSTSTSLSTSNSEST	1152
Qy	1309	PE-NDTKSLTVKES-----AIRPFG--DVIMEDFNERNSETK-----S	1346
Db	1153	TSMSDSTSLSTSESDSTSDSTSDISPAISESTSI:SESNSTSDSESKASAPLS	1212
Qy	1347	HLSSDAE-----GNTRDLETLPTPKESDSTOTTTPASCPESNVQVEDMEIET	1399
Db	1213	ESLESTSESTSESLSGSTSDSTSLSDNSESGSTSTSL-SNSTSGSASISTSTSGSAST	1271
Qy	1400	SEVKVTISPITSEESNLSNDFIDNGULPINKNENVGE-----SKRKT	1444
Db	1272	STVK-----SESVSTSLSTSTSLSDSTSLSTSLSDSTSGKSNLSASMSSTSDSISTRKS	1328
Qy	1445	VITEVTMTSTVATESK--TVIKVKEGDK-----QTVVSTENCASKTVT-	1487
Db	1329	ESLSASTSLSGSTSESESGSTSSSESKSDSTSMLSMSQSTSGSTSVSTSESLDSTSTS	1388
Qy	1488	-----TTTTVTAKLSTPTGGSDVLIISVKEQSKTVVTTVTVD	1525
Db	1389	LSLSASMNQGVDSNASQASASTSTSTSESDSQSTSSYTSQSTSESTSTSTSLSDS	1448
Qy	1526	L-----TTTGGTLVTSMTVSKYSTDKVKLMKFRPKKTRSGTALPYRKVFVKSTKKS	1580
Db	1449	TSISKSTSGSGSTSTASLGSSESDSQSTSTSESKSESTSTSLSDSTSTSMGSGAS	1508
Qy	1581	IFVLPNDDLKLLARKGIRVFPYFNNAKPALDIWPPSPRPTFGITWYRLQTVKSLAG	1640
Db	1509	TSTL-----LNSASASESDSSSTSLSDSTASMQSSESDS-----QSTSTSLSN	1553
Qy	1641	VSLMLRLNLASLRWDDMAAKVPGGSGSTRFETSEITTEILIKRRDVGCPYIREYCI	1700
Db	1554	SQSTSTSI:RMSTIASSEVSESTSESGSTSESTSESTSTSL-----	1595
Qy	1701	KIICPIGPETPKPTPPQKGLRSSALRPKRPTKQTPGVIIETWVAEELELWEIRA	1760
Db	1596	-----SDSQSTSTSESTASGASASTSTSTSDSRSTASTSTSMTSLD-----	1637
Qy	1761	FAERVEKEQAQVEQAKKLEQOKPVTIATSTSTSTSTSTSTSPAQKVMVAPISGVT	1820
Db	1638	-----SQMSLSLSTSTSVSDSTSLSDSVSDSTSDSTSTSTSGMSASISLSDSTS	1688
Qy	1821	TGTKM--VLITKVGSPATVTFQONKPHOTFATVWVGOGSNCGVVOQKVLGIIPSTG	1878

Query Match		2.8%;	Score 401.5;	DB 10;	Length 1367;		
Best Local Similarity		21.2%;	Pred. No. 3.4e-12;				
Matches 289;		Conservative 168;	Mismatches 564;	Indels 339;	Gaps 56;		
QY	1135	DPEG---	KLGDSE-----	SNSTLENSSDTVSIQDSS	SEEDMIVQNSNISISQFRT 1182		
DB	191	DPCGYWIDC	NNCGTKSGTS	TTTSSESTTS	STSESS-----TTTSSTSESST 243		
QY	1183	REQDVEVLE	PLKCLVSGESTGC	NC--EDRLPVKTEANG	KKPKPSQOKKLEERP	VNKCSQDI 1240	
DB	244	TSSTSESS	STSTTAPATPTT	SCTKEKPTPTT	SCTKEKPT-----PHDHTTPTCT 296		
QY	1241	KLKNTTDK	NNENRESEK	KGORTSTFQINGDKN	PKIYILKGECLKEISE	SVSVNVEPK 1300	
DB	297	KKKTTT	SKTKTKTKTT	TPVPTPSSSTTE	-----SSGAPVTPPSS	STTESSAPVTSSTTE-- 350	
QY	1301	VNNKII	PENDIKSL	TVKESAIRPFINGD	VIMEDFENERNSSE	TKSHLLSSSDAEGNYRD 1360	
DB	351	--SSS	APVP---TPSS	STTESSAPVTS	-----STTESSAPVTS	STTESSAPVTSSTTESSAP 394	
QY	1361	SLETLPT	SKESDS---TQ	TTTPSACPSNSVNQ	VEDMEIETSEV	KKVTSPTISE--- 1413	
DB	395	VPTPSS	STTESSAPVTS	STTESSAPVTS	STTESSAPV-TS	STTESSAPVTSSTTES 453	
QY	1414	-----EESN	--LSNDFID	ENGLPINK	NNVNGESK	RKTVITEVTMTST-VA 1457	
DB	454	SSAPV	PTPSSSTTESS	APVTSTTESS	APVPPTPSS	STTESSAPVTSTTESSAPV 513	
QY	1458	TESKTVI	KVEGDKQ	TVVSSTENCAK	STVTTTTTVTKL	STPSTGGSDIISVKEQSKTV 1517	
DB	514	TPSS	STTESSAPAPT	PTPSSSTTESS	APVTSTTESS	APVTPSS-----STTESSSTP 568	
QY	1518	VTTTVD	SLTTGGTLV	TSMTVSKB--	YSTKDKVKLM	KFSRPKTRSGTALPSYKRVTK 1575	
DB	569	VTSS	STTESSAPVPT	PTPSSSTTESS	APVPPTPSS	STTESSAPAPTPTPSSSTTESSAPVTS 628	
QY	1576	STKKSIF	VLNDLKL	ARKGIREVPVFN	YNAPALDI	WPYPSRPTFGITWRYKLQTV 1635	
DB	629	STTESSA	-----	-----	-----	VPPTPSS-----TTE 647	
QY	1636	KSLAGV	SLMLRL	MLASLRWD	MAAKVP	PGGSTRTETSEITIT-----TEIKR 1685	
DB	648	SSSAPV	PTP-----	SSSTTESS	APVPTPSS	ST-TESSAPVTS	STTESSAPVTSSTTE 701
QY	1686	RDVGP	GIRFEX	YCIKIKPI	GPV-----ETP	KETPTPQKGLRSSAL----- 1728	
DB	702	SSSAPV	PTPSSSTTESS	APVPPTPSS	STTESSAPVPTPSS	STTESSAPVTSSTTESS 761	
QY	1729	----	RPKR	PTPKOT	GPVIIETWVA	EELEWIRAFERKEKAQAVEQAKKLEBQ 1784	
DB	762	APVPTP	SSSTTESSAPVPTPSS	STTESSAPVPTPSS	STTESSAPVPTPSS	STSSNITSS 821	
QY	1785	KP--	TVIAT	STTS-----	PTSS	TTTISPQKVMVAPISGVTTGKMLTKVGS-- 1833	
DB	822	APSST	PTPSSSTTESSVPVPTPSS	STTESS-----	APVSS	STTESSVAPVPTPSSSN 874	
QY	1834	----	PATV	TQOKNHFOT	ATWVKQGS	GVVQOVKVLG-----IIPS----- 1875	
DB	875	ITSSAP	SSIPSSSTTESFST	-GTV	PTPSSSKYPGSQ	TETSVSTTETIVPTKTTSVTT 933	
QY	1876	-----	STGTSQ--	QFTTSQ	PRATVTRPNT	ISGSGTTSNSQVITGQIRPGM 1922	
DB	934	PSTTT	TTTTCVSTG	TSNAGETSGC	SPKTVTTV-PT	TTTTTSVTSTTTT-----T 985	
QY	1923	TVIR	TPLO--	---QST	LG---KAI	IRT-PVMVOPGAQOVMTQIIRGQVPVSTAVS-----AP 1970	
DB	986	TVCST	GTNSAGETSGC	SPKTIPTTV	PCSTSPSETASE	STTSTPTPTVTVSVTVTE 1045	
QY	1971	NTVSS	TPQKSL	TATSTNTSQSS--	ASQPP	RRQOGQV-KLTMQLTQLTQHGNGOGLT 2027	
DB	1046	YSTST	KPGGE-IT	TTFTVTKN	PTTYLTAT	IAPTPTSVTTVTFNPTTTTITTTVCSTGTNSA-- 1102	
QY	2028	VVIQ	OGQTTQ	LQIL	IQGVTVL---	PGPQQLMQAAMPNGTVQRFLEPTLATATAS 2083	

DB	1103	-----	GETTSGCS--	PKTVTTT	PCSTGTGEYTT	EA-----TTLVTTA--- 1133	
QY	2084	TTTTTT	VTTAAAGTGEOR	QSKLSPQ	MQVHODKTLPP	APQAQSPVPAQAQSPQARSARPOQ 2143	
DB	1139	-VTTT	VTTTSTSTG	NSAGKTTTG	---YTTK	SVPTTYVTTLAPS----- 1178	
QY	2144	TOQSPA	QPEVQTOPEVQ	TQTTVSS	HVPSEAQPT	HAQSSKPPQVAAQSQPSNVQGSPPVR 2203	
DB	1179	----	APVT	PATNAVPTTIT	TTECSAAT	NAAGETTSVCSAKTIVSSASAGENTAPSAT--- 1231	
QY	2204	VQSPSQ	TRTRPST--	PSQLSP	QGSQOVQTT	---TSQPI	PIQHTSLOIQSQSQSQSQPOV 2258
DB	1232	--TPV	TAT-PTTV	ITTES	SVGTNSAGET	TTTGYTKS	IPTTYITL-IFGSGAKNYETV 1287
QY	2259	QSS	TOTLSSGQTL	NVSVSS	PSRPOL	IQPOQVIAV	PQLQQOVQLSQIOGVVVAQIQ 2318
DB	1288	ATA	T-----	NPIS	IKTTS-----	-----QLAT	TASASSVAPVV 1315
QY	2319	AQSGV	PQOIKLQ	LP	IQIOQSS	AVQTHQ	IONV-VTVQAAS 2357
DB	1316	TS	PS-----	LTG	PLQ	SASGSAVAT	YSVPSSISTVQGAA 1348

RESULT 11

US-09-815-242-12713

; Sequence 12713, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12713

; LENGTH: 2344

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-12713

Query Match

2.8%;

Score 401;

DB 9;

Length 2344;

Best Local Similarity

18.0%;

Pred. No. 7.3e-12;

Matches 336;

Conservative 271;

Mismatches 836;

Indels 424;

Gaps 57;

QY

797

EGT

KNM

DM

NB

DES

DK

R

CS

RP

KK

IK

IE

PD

SE

K

DE

V

K

S

Q

S

V

S

T

S

G

S

T

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

G

S

T

S

Query Match 2.8%; Score 401; DB 9; Length 2344;

Best Local Similarity 18.0%; Pred. No. 7.3e-12;

Matches 336; Conservative 271; Mismatches 836; Indels 424; Gaps 57;

Qy 797 EGTNNMDENDESKRCSRPKKIKTEPPSEKDEVKSDAAGADQNMDSKITEKK 856

Db 747 EVTRNSMDSYSTSGSTQQSQSVSTSKADQSASTSGS----- 786

Qy 857 DQVKELDSDSDRCKEPPMEVDDDMKTESHVNCQESSQV-----DVVNVSEGFHLRT 910

Db 1478 ---VCTDKPTTTHKTKIACRSQPD-----PVDPTSSKPSQSR 1513
QY 1389 VNQVEDMEITSEVKKVTSS-----PITSEESNLSNDIDENGLPINK---NENV 1436
Db 1514 SLRKVDVEEFFALRKRTSPSAGKAMHTPKPAVSGEKNYA--FM--GTPVQKLDLTENL 1568
QY 1437 NGESKRKTVITEVTTSTVA-----TESKTVIKVKGDKQTVVSTENCAS 1484
Db 1569 TSKRRRLQTPKEKAQALEDLAGKELFQTRGHTSEESMT---NDKTAACKVSSQPDLDKN 1625
QY 1485 TVTTTTTVTKLSTPSTGGSDVSIYSKEQSTVTVTTTVDLSLTTTGGTLTMTSVSKEYS 1544
Db 1626 PASSKRLKTSLG--KVGVEKELLAVGKLTQTSGETTHTTPTGDKSKMAFESPQOI 1683
QY 1545 TRQVKUMKFSRKPRTSGTA-----LPSYK--FVTKSHKSFIVLPND 1588
Db 1684 LDSAASLTGSKRLQRTPKGSEVPDLAGFIELFQTPSHTKESMTNEKTTKVSYRASQPD 1743
QY 1589 L-----KKLARKGGIREVPVFNNAKPALEDIWPVSPRPTFGITWRVRLQTVKS 1637
Db 1744 LVDTPTSSKQPKRSLRKADTEE-----EELAFKQTPSNAG-----KAMHTPKP 1787
QY 1638 LAGVSLMLR--LLWASLRWDDMAAKVPPGGSTRTTSETTEITTEILKRRDVGPIGIR-- 1694
Db 1788 AVGEKDIINTFLGTPVQKLDQGNLP--GSNRRLOTRKEKAQALELT-----GPREL 1838
QY 1695 -----FEYCIRKIIC--PIGVP-ETPKETPTPQKGLSS-----ALRPKRP 1733
Db 1839 FQTPCTDNPTTDEKTKTKIILCKSPQADPTNTKORPKRSLKKADVEEFLAFKRLTP 1898
QY 1734 -----ETPKQTPGVILETVAERELELWEIRAF-----AERVE-----KE 1768
Db 1899 SAGKAMHTPKAA-----VGEK-----DINTFVGTPEVKLDLLGNLPGSKRRPQTPKE 1946
QY 1769 KAQAQVEOAKRLRLEQKPTVIASTTSTSTTISPAOKVMVAPISGVSVTGTKMVL 1828
Db 1947 KAKALEDLAGKELFQTPGHTSEESMTDDKITEVCSKSPQDPVKTP-----TSSQRLK 2000
QY 1829 TKVGSPTATVFOQKNKHQIFATVWVQKQSGNSGVVQKVL--GIIPSTSTGTSQOFTS 1886
Db 2001 ISLKG-----VGKVEVLVPVKGKLTQTSKGKTQT--- 2028
QY 1887 FQPRATVTRPNTSGSGGT-----TNSQVI-----TGQIRPGMTVIRTP-----L 1929
Db 2029 -----HRETAGDKSIFAKESAKQMLDPANVTGMRWP-----RTPKEAQL 2073
QY 1930 QOSTLGRKAITPVMOPGAPQVMTQIIRGQPVSTAVSAPNTVSSTPGQKSLTSATSTS 1989
Db 2074 EDLAGFELFQTPDHTTEESTTDDTKTKI-----ACKSPPPESMDTPTST- 2117
QY 1990 NIQSSASOPPRPQOGQVKLT--MAOLTQLTQGHGNGQLTVVIOGQOTTGQLQIP--- 2044
Db 2118 -----RRRKTPLGKRDIIVEELSALKQLQ-----THTDKVPQDE 2153
QY 2045 -QGVTVLPGQQLM-----QAAMPNGTVQ-----RFLF--TPLATTATAS 2083
Db 2154 DGINVFPRETAQKLDPAASVTGSKRPRTPKGAQPLEDLAGLKELFQTPICTDKPTTH 2213
QY 2084 TTTTIVSTTA-----AGTGE--QRQSKLS--FOMQVHQDKTLPAPQSSSVGPAKAQPTA- 2134
Db 2214 EKTTKIACRSQPDVPVGTPPTIFKPSKRSRLKADVEESLALRKRTPSVGMADTPPKPAG 2273
QY 2135 -----QPSARPOQTPQSPQAPEVOTQPEVOTQTTVSS-----HVP 2171
Db 2274 GDEKDKAFMGTVPVQKLDLPNLPKSKRWPTPRE-----KAQALEDLAGKELFQTP 2326
QY 2172 SEAOPTHAGSKPQVAAQSPQSNVQSPVRSQSPQTRTPS----- 2215
Db 2327 GTDKPT--TDEKTKTKIACKS-PQ-----PDPVDTTASTKQRPKRNLRKADVEEFLALR 2377
QY 2216 ---TPS-----QLSPGQSQVQTTSTPIPIQPHFTSL-QIP-SQOQOSQP----- 2256
|||

Db 2378 KRTPSAGKAMTTPKPAVSDKKNITFETPVQKLDLLGNLPGSKRQOPTPKKAEALDEL 2437
QY 2257 -----QVQSSPOTLLSSGOTLNOVSVSPSRPQLQIQOPQPOVIAPV---QLQOQVQV 2305
Db 2438 VGFKELFQTPGHTSEESMTDDKITEVCSKSPQSPESFKTSRSSKORLALPLVKVDMKEPLA 2497
QY 2306 LSOI--QSQVVAQIQAOQSGVPOQIKL--OLPIQIQOQSSAVQTHQTONVVTVQAAVQEQ 2361
Db 2498 VSKLRTSGTETTQHTPTGDSKSIKAFKESPKQILDPA-----SVTGSRRQLRTR 2549
QY 2362 LQRVQOLRDOQKKKKQOIIEIKREHTLQASNQSEI-----IQ 2398
Db 2550 KEKARALEDLVDFKELFSAPGHTSEESMTIDKNTKIPCKSPPPPELTDTATSTKCKPKTRLR 2609
QY 2399 KOVVMKHNAVIEHLKQKSMT-----PAEREENQRMIVCNQVMKYILDKIDKEEKOAAK 2452
Db 2610 KEVKEELSAV-ERLTQTSQSTHTHKEPASGDEGIKVL-----KORAK 2651
QY 2453 KRKEESVEQKRSKQNA-----TKLSALLFKHKEQLRA----- 2485
Db 2652 KKPNEVEEESRRRPRAPEKEAQLPDLAGFTELSETSGHTOESLTAGKATKIPCSPPL 2711
QY 2486 EILKKRALLDKDLQIEVQEEELKRDLKIKKEKDLMLAQ----- 2523
Db 2712 EVVDTTASTKRLHRLTRVQK-----VQKPEPSAVKFTQISGETTDADKEPAGEDKGIKAL 2766
QY 2524 ---ATAVAACPVPVPLPAPPAPPSP-----PPGVQHTGLLSTPTLPLVASQKR 2572
Db 2767 KESAKQTPPAASVTGSRRRPRAPRESAQAIEDLAGFKDPAAGHTESMTDDKTTKIPCK 2826
QY 2573 KRBEEDSSKSKKK-----KMSTTSKETKKDKTKLYCICKTKTPVDES 2614
Db 2827 SSPELEDATSSRRRPRTAQRKVEKBELLAVGKLTQTSGETTHTD-----KEPVGEG 2879
QY 2615 K 2615
Db 2880 K 2880

RESULT 10

US-09-801-368-108
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
; LENGTH: 1367
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108


```

; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34248
; LENGTH: 2665
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034555.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
; OTHER INFORMATION: EST_HUMAN HIT: AU117052.1, EVALUOE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUOE 3.00e-10
US-09-864-761-34248
```

```

Query Match          2.9%; Score 416.5; DB 9; Length 2665;
Best Local Similarity 18.5%; Pred. No. 1.4e-12;
Matches 544; Conservative 422; Mismatches 1145; Indels 825; Gaps 132;

Qy      7  EEDGDAAETQDSEDE-----EDMEEDDDSDSDYPEEMEDDDDDASYCTESSFRSHS 59
Db      170 KNETDKERTFDPVERERRLIRKEVKKDKTKQ----- 205

Qy      60  TYSSTPGRRKPRVHRPSPILE-----EKDIPPLEFPKS-----SEDLMVPNEHI 104
Db      206 -----KRKGKVVHSPSSOSSETQENEREQSP-EKPRSCNKLRSERADKEGIAKNRLEL 257

Qy      105  MNVTAIYVLNFGTVLRLSPREFEDCAALVSQEOCTLMAEMHVLLKAVLREEDTSNT 164
Db      256  MPCVVLTATVKKKEGKVIDHTV-----EKLKAKLDNDTVKSSALDQKLQVSQT 305

Qy      165  TFGPADLK--DSVNSTLYFIDGM-TWPEVLRVYCESD---KEYHHVLPYQBAEDYPYGPV 218
Db      306  EPKASDLSKLESVRMKVPKESGLSSHVEV---EKEGRKARKHLKPEQPAD---CVS 357
```

```

Qy      219  ENKIKVQLVLDQOFLTNTIAREELMSE---GVIQYDDHCRVCHVKHGLDLCCCTCSAVYHL 275
Db      358  AVDLKLEAKRRFADSNLKAQKQKPEVKKSSPEMED-ARVLSKKQPDVSSREVILLREG 416

Qy      276  ECVKPL--EEVPEDEWQCEVCVAHKVPGVTDCAVIAIQNKPKYIRHEPIGYDRSRKYWF 333
Db      417  EAERKPVKKEILKRESKIKLDRLNTVASPKDC-QELAS-----ISVSGSGSPSSD 466

Qy      334  LNRRLL--ITEEDTENENEKKIWIYSTKVOLAEELIDCLDKDYWEAELCKKILEM---REEI 388
Db      467  LQARLGELAGESVENQEVQSKKPIPSKPLKQL-----QVLDLDOGPREDV 512

Qy      389  HR-HMDITEDLTNKAQSNKSFLLAAANEELIESIRAKG-DIDNVKS---PETEKDKNE 443
Db      513  RNYCYSLRDETPEKSGQEKSHSVNTEEKI-----GIDIDHTQSYRKQMEQSRKQ 564

Qy      444  TENDSKDAE-----KNREFEFEOQL-----EKSDDKTPTDDPEQCK-SEVGDF----- 486
Db      565  MEMEIAKSEKFGSPKDKVDYEYERRSLVHEVGKPPQDVTDSPSPSKKRRMDHVDICTKR 624

Qy      487  -KSEKNGELSE-----SPGAGKA-----SGSTRIITRLRNP--DSKLSQLKSQ 529
Db      625  ERNYSRQISEDSERTGGSPSYRHGSFHEDEDPICSPRLLSVKGSPKVDKVLPSNIT 684

Qy      530  VAAAHAHEANKLFKEGKEVL-----VYNSQGEISRLST-----K 562
Db      685  VREESLKFNPYDSSRREQMADMAKIKLSVLNSELNRMWDSQKODAGREFDVSPNIIK 744

Qy      563  KEVIMKGNINNY-----FKLGQEGKYRVYHNOYSTNSFALNKH-----QHRED--- 605
Db      745  RDSLKRKVRDLPEGVEPVSDEDEG-----HKSHSPASALYESSRLSFLLRDREKLR 800

Qy      606  HDKR-----RHLAHHKFL-----TPAGEFKWNGSVHGSKVLTITLRLTITOLENNIPSS 655
Db      801  RDERLSSLERKNKFYFALDKTTP-----DTKALLERAKSLSSREEN---WS 846

Qy      656  FLHPNNASHRANWIKAVOMCSKPREFALALAILCAVKKPVVMLPIWREFLGHTLRLHMTS 715
Db      847  FL--DWDSRFANFRN-----NKDKE-----KVDSAPRPI---PSW--YMKKKIR--- 884

Qy      716  IEREKEKVKKKEKKQBEETMQATWVKYTFPVKHQVQKQGEYRVGTGGWSWISKT 775
Db      885  TDSEGMKMDKKEHKEEEQEQE-----LHSSIFEQDSKRLQHLERK---BEDSDFISGRI 908

Qy      776  HVYRFVKPLPGNTNVNYRKSLEGTNNMDENMDESRRKCSRPPKTIKIPDSE--KDEV 833
Db      909  FASRF-----LHSSIFEQDSKRLQHLERK---BEDSDFISGRI 943

Qy      834  KGSAAKAGADQNE-----MDISKITEK---KDQDVKELLDSDS---DKPC 872
Db      944  YGKQTSSEGANSTTDSIQEPVVLPHSRFMELTRMOQKEKQDQKPKVEKQEDTENHPKTP 1003

Qy      873  KEPEMEYDDDMKTESHVNCQESSQVDVNVVSEGFLRTSYKKTKSKSLDGLL---ERRI 929
Db      1004  ESAPENKDSCLKTPPSVG---PPSVTVVTLESA---PSALEKITGDKTVEAPLVTSEKTV 1057

Qy      930  KQFTLEK-----ORLEKIKLEGIGIKGIGTSTNSKNLSSESPVIYAKGCGQ 977
Db      1058  EPATVSEAKPASEPAPAPVEQLBQVDLPPG-----ADPQKEAAMP--AGVERGSS 1107

Qy      978  SDSNRQSGSPNANDQPEDLIQCGSQSDSVSLRMSDPSTHTNKLYPKDRVLDVDSIRSPE 1037
Db      1108  GD-----QPPYLDLAKPPTPGASFSQAESNVDPDPDSTQPLSKPAQSEANEPKAKPD 1161

Qy      1038  TKCPKQNSIENDIEEKVSDLASRQGEPTKSKTKGNDFIDDSKILASADDITGLTCKNKKP 1097
Db      1162  ATA---DAEPDANQK-AEAAPESQPPA-----SEDLEVDPVPAV-----KDKKP 1201

Qy      1098  LIQESTST-IVSSSKSALHSVPKSTNDRATPLSRAMDPEGLKGCDSNSTLENSDT 1156
Db      1202  NKSRKSTPTVQAAAVSIVEKPVTRKSRIDREKLKRSNPRG-----EAKLLELKMEA 1255
```


; ORGANISM: Homo sapiens
US-09-854-133-425

Query Match 2.9%; Score 418; DB 10; Length 4019;
Best Local Similarity 17.4%; Pred. No. 2e-12;
Matches 492; Conservative 320; Mismatches 995; Indels 1018; Gaps 103;

QY	254	CRVCHKL---GDLLCETCSAVYHLECBVCPPLBEVPEDEWCEOCVCAHKVPGVTDCAVAEI	310
Db	65	CEACGRATDPGRLLLLCCDDCDSYHTYCLDPLQTPVPGGWKCKWCVMCRHCGATSAGLRC	124
QY	311	QKNKPYIRHEPIGYDRSRRRYWFLNRRLLITEEDTENENEKKIWIYSTKVOLAEILDLCLK	370
Db	125	EWQNNYTQCAPCA-SLSSCPVCYRNYR---EEDL-----ILOCRQC	161
QY	371	DYWEAEUCKILEMREBEIHRHMDITEDTLNKAAGSNKSFALAAH-----EILBSIR	422
Db	162	DRMHAVCQNL-NTEBEVENVADIGFCS-----MCRPYMPASNPVSPDCCESSLVAQIV	215
QY	423	AKKGIDINVKSPPE-----TEKDKNETENDSKDAEKNREEFEDQSLSEKSDDKT-----	471
Db	216	TKVKELDPPTYQTQDGVCLTSEGMTQLOSLTVVPRKRKPKLKLIINQNSVAVLQTP	275
QY	472	PD-----DDPEQKSEVDFKSEKSNSELSPEGAGKASGSTRITITLRNPD	520
Db	276	PDIQSEHSDGEMDSDREGELMCDCKSE--SSPEREAVDDETKGEVTDGVKKRKRPP--	332
QY	521	KLSQLKSQVAAAAHEANKLFPKGEKGVLVNSQGEISRLSTKKREVMK---GNINNYFKL	577
Db	333	-----YRGIGGFMYQRKSRTOGGKTKRSVIRKDSGSGSISEQLPC	372
QY	578	GOBEKYRVYHNOYSTNSFALNKHQHRRDHDKRRHLAHKFLCTPAGEFKWNGSVHSGKVL	637
Db	373	RDG-----	376
QY	638	ISTLRLLITOLENNIPSSFLHPNWSHRANWIKAVQMSKPRFALALALECAKVPVM	697
Db	377	-----	376
QY	698	LPIWREFLHTRLRHMTSIEREEKEVKKKEKQEE--EET---MQOATWVKYTFPVKH	751
Db	377	--WSEQLPDLVDESYSV-TESTEKIKKRYRKRKNKLEETFPAYLQEAFFGKDLLTSTR	432
QY	752	QVWKQKGEERYVTGYGGWSHISKTHYRVPKPLPGNTNVNYSKLEGTKNMNDENDESD	811
Db	433	Q-----SKISLDNSEDGAQLLYKTNN-----TGFDPSPSLD--P	465
QY	812	KRKCRRSPKKIKIEPDESEKDEVAAGSAAKADONEMDISKITEKKQODVVELLSDSDKP	871
Db	466	LLSSSSAPTK-----SGTHGAPDPLADISEVL-NTDDDLIGIISDDLAKS	510
QY	872	CKEPM-EVDDDMKMTSHVNCQSSQDVNVNVEGPHLRTSYKKKTKSKLDGLLERRIK	930
Db	511	VHSDTIGVPTDPSLPPQPNVNOSS-----RPLSEEQILDGIL-----	547
QY	931	QFTLEEKORLEKTKLEGGIKG-IGKTTNSTSSKNLSPEVITTKAKEGQSDSMRQEQSP--	987
Db	548	-----SPELDKVVTDGAILGLKYKIPELCGKDVEDLFTAVLSPANTQPTPLPQPPPTQ	601
QY	988	-----NANNQDQEDLIQCGSQSDSSVLRMSDP-----SHTNKNLYPKDRVLUDDYSIR	1034
Db	602	LLPIHNOQAFSRPMLNGLIGSSPHLPNLSLPPGSGGLGTFSAITAOSSVPDAR--DKNSAF	659
QY	1035	SPETKCPQN-----SIENDIEEKVSDLASRQEPTKSKYTKGNDFFIDDSKLASADDIG	1088
Db	660	NPMASDPNNNSWTSSAPTEGE-----NDTMSNAQRSTLKWEK-----EALGEMATVA	707
QY	1089	TLICKN-KKPLIOBESPTIVSSSKSALHSSVPKSTNDTRDTPLSRAMDFEKGLCD--SE	1145
Db	708	PVLYTNIINFNLKEEPPDWITRVKQ-IAKLWRKASSOERAPYVQKARDNRAALRINKVQM	766
QY	1146	NSNTESSNPTVSIQDSS-----EDMIVQNSNESISEFRTREQ-----DV	1187

[illegible]


```
QY 1923 -----TWIRTPLOQSTLGAIRTPVMVQPCAP-----QQVMTQ-----IIRGPVS 1964
Db 2809 NTSEGVLLSYSGQTEGPORISAKISQIIPPASAMDIFFQOSVSKSQVQKPDVSTASQPPS 2868
QY 1965 TAVSAPNTVSTPGOKSITSATSN-----IQS-SASQPPRPOQOGQVKLTMAQLTQLTQOG 2019
Db 2869 KGQOAPAGYANVATHSLVLTQATYNASPVYSSVKADRPSELEKPEPIHLSVS---TPVTOG 2926
QY 2020 HGGNOGLTVVJQOGQTT-----GQLLIPQGVTV---LPGP-----GOOLMQAA----- 2061
Db 2927 G-----TVKLTQGTGTPPVVLVHNLVLTPTSIYVTTNKKLADPVTLKIKETKVLQPANLGS 2980
QY 2062 -----MPNG-----TVQRELFTEL----- 2075
Db 2981 TLTPHHPALPSKLPTEVNVHPSGSPDAPKTSVSHLAAKLDHANSRPSGPGSPFFRAS 3040
QY 2076 --ATTATTASTTTTSTTAAGTGEORQSKLSQMVHQDKTLPQAQSS-----SVG 2125
Db 3041 HPSSTASTALSTNATVMLAAGIPVPQFISSIHPSQSV---IMPPHSITQTVSLSHLSQG 3096
QY 2126 PAKAQPTAQBP---SARPOQOTQPSQAQPEVQTOPEVQTOPTVSSHVPSQAQTHAQS 2182
Db 3097 EVRMTPTLPSITYSIRPEALHSPRAPLQPO---QIEVRAPQRAST-----POPAPAG 3146
QY 2183 KPOVAAQSQPSNVQOGSPV-RVQSPSOT----- 2210
Db 3147 VPALASQHPPEEYHYHLVARATAPVOSEVLVMQSEYRLHPYTVPRDRVIMVHPHTAV 3206
QY 2211 -----RIRPSTPSQLSPGQOS-----QVQTTTSQPIPI-----QPHTS 2243
Db 3207 SEQPRAADGVVVPFASKAPQPCKEAKTPDAAKAPTTPAPVPVPLPAPAPAPHGE 3266
QY 2244 LQI-----PS---QQQOSQOVQSSQTOL---SSQOTLNOV----- 2274
Db 3267 ARILTVPNSLOGLPTPPVVVTHGVQIVHSSGELFQERYGDIRTYHPPAQLTHRTQFP 3326
QY 2275 ---SVSSPSR-----POLQIQOPQOPQVIAPV-----OLQOQVLSQIQS 2311
Db 3327 AASSVGLPSRTKTAQAQPPPEGEPLQPPQPVOSTQPAQAPCPQPSQLQGPQPPSKMP 3386
QY 2312 QVVAQIQAOQSGVPPQIKLQLP-----IQQQSSA 2341
Db 3387 QVSQEAQTGTGVQEP---RLPAGPANRPPPHPTQVQORAQA 3424
```

RESULT 4

US-09-738-973-425

; Sequence 425, Application US/09738973

; Patent No. US20020110563A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Henderson, Robert A.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Fling, Steven P.

; APPLICANT: Mohamath, Raodoh

; APPLICANT: Algate, Paul A.

; APPLICANT: Secrist, Heather

; APPLICANT: Indrias, Carol Yoseph

; APPLICANT: Benson, Darin R.

; APPLICANT: Elliot, Mark

; APPLICANT: Mannion, Jane

; APPLICANT: Kalos, Michael D.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

; FILE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.475C9

; CURRENT APPLICATION NUMBER: US/09/738,973

; CURRENT FILING DATE: 2000-12-14

; NUMBER OF SEQ ID NOS: 587

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 425

; LENGTH: 4019

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-738-973-425

Query Match 2.9%; Score 418; DB 10; Length 4019;

Best Local Similarity 17.4%; Pred. No. 2e-12;

Matches 492; Conservative 320; Mismatches 995; Indels 1018; Gaps 103;

```
QY 254 CRVCHKL---GDLCCETCSAVYHLECVKPPLEVPEDWQCEVCVAHKVPGVTDCAVEI 310
Db 65 CEACGKATDGRLLCLDDCDSYHTYCLDPLQTPVPGWKCKWCVMCRCGATSAGLRC 124
QY 311 QKNKPYIRHPIGYDRRRKYWFLNRLITBEDTENENKIKWYISTYKVLQAEILDCLDK 370
Db 125 EQWNNYTCAPCA-SLSCPCVCYRNYR---EEDL-----LQCRQC 161
QY 371 DYWAELCKLLEEMREIHRMDITEDLTNKARSNKSFLLAAN-----EELFSIR 422
Db 162 DRNMHAVCQNL-NTEEEVENVADIGFDCS-----MCRPYMPASNPSPDCCESLSVAQIV 215
QY 423 AKKGDIQNVKSPEE-----TEKDKNETENDSKAEKNREEFEDQSLEKSDDKT----- 471
Db 216 TKVKELDPKPYTQDGVCLTESGMTQLOSLTVTPRRKRSKPKLUKLIINONSVAVLQTP 275
QY 472 PD-----DDPQKGSEVDFKSEKSNGLSESPGAGKAGSASTRIITRLRNPDS 520
Db 276 PDIOSEHSRQCEMDDSREGLMCDCKSE-SSPERAVDDDETKGVEGTDGVKKRKP-- 332
QY 521 KLSQLKSQVAAAHAANKLFKEGKEVLVNSQGEISRLSTKKEVIMK---GNINNYFKL 577
Db 333 -----YRPGIGFMVRQSRGTGOKTKRSVIRKDKSSGISSEQLPC 372
QY 578 GQEGKYRVYHNOYSTNSFALNKHQHRHDHKKRHLAHKFLCTPAGEFKWNGSVHGSKVL 637
Db 373 RDDG----- 376
QY 638 ISTLRLTITOLENNIPSFLLHPNASHRANWIKAVQMCSPREFALALALECAVPPVM 697
Db 377 ----- 376
QY 698 LPWREFLGHTRLRHMTSIEREEKEVKKKEKQEE---EET-----MQQATWVKVTFPVKH 751
Db 377 ---WSEQLPDTLVDESVSU-TESTEKKIKRYKRRKLEETFPAYLQEAFFGLDLSR 432
QY 752 QVWKQKGEEYVTCYGGWSMISKTHVYFVPKLPNTNVNRYKSLGTEKNNMDENMDES 811
Db 433 Q-----SKISLDNLSEDAQLLYKTNMN-----TGLDPSLD--P 465
QY 812 KRKCSRSPKKIKIRPDSEKDEVGSDAAKADQONEMDISKITEKKDQDVKELLDSDDKP 871
Db 466 LLSSSSAPTK-----SGTHGPADDDPLADISEVL-NTDDDLILGIIISDDLAKS 510
QY 872 CKEEPM-EVDDDMKTESHVNCQESSQVDVNVVSEGFLHRTSYKKTKSKLDGLLERRIK 930
Db 511 VDHSDIGPVTDDPSLPQPNVQSS-----RPLSEQLDGL----- 547
QY 931 QFTLEEKORLEKIKLEGGIKG-ICKTSTNSKNLSESPVITKAKEGQCSQSMRQEQSP-- 987
Db 548 -----SPELDKMVTGAILGLKLYKIPELGCKQVEDLFTAVLSPANTQPTPLPQPPPTQ 601
QY 988 -----NANNDQPEDLIQCSQSOSSSVLMSDP-----SHTTNKLYPKDRVLDDVSIR 1034
Db 602 LLPHNODAFSRWPLMNLGLTSSPHLPHNSLPQSGSLGTFSIAIAQSSYPDAR--DKNSAF 659
QY 1035 SPETKCPQN-----SIENDIEKVSQSLASRGQEPKSKTKGNDFFIDSKLASADDIG 1088
Db 660 NPMASDPNNSWTSSAPTVEGE-----NDTMSNAORSTLKWEK-----BEALGEMATVA 707
QY 1089 TLICKN-KKPLIQEESTIVSSSKSALHSSVPKSTNDRDATPLSRAMDPECKLGD--SE 1145
Db 708 PVLVTNINFPNLKEEFPDWTTRVKQ-IAKLWRKASSQERAPYQVOKARDNRAALINKVQM 766
QY 1146 SNSTLENSDTSVSIQDSSE-----EDMIVONSNEISIEQFRTREQ-----DV 1187
Db 767 SNDSMKRQQQODSIDPSSRIDSELFKPLKQRESEH-EQEWKFRQQRKQSKQAKTEAT 825
```

875	MPCVVLTRYKEGKVIDHTPV-----EKLKAKLDNDTVKSSALDQKLOVSQT	922	QY	1069	---TKGNDFFIDDSKLASADDIGTLCKNKKPLIQEESDTIVSSSKSALHSSVPKSTNDR	1123
165	TFGPADLK--DSVNSTLYFIDGM-TWPEVLVRYCESD---KEYHHVLPLYQEAEDYPYGPV	218	QY	1838	KPVTRKSE-RIDREKLKRSNPRGEAQKLELAMEAEKTRTASKNSAADLHPE-----	1891
923	EPAKSDLSKLESVRKVPKKEGLSSHVEV---EKEGRLKARKHLKPEQAD-----GVS	974	Db	1136	DATPLSR-----AMDFEGLKGCDSSESNTLSENSDTV-----SIODSEEDMIVON	1171
219	ENKIKVLQFLVDQFLTTHIAREELMSE---GVIOYDHCRCVCHKLDGLLCCETCSAVYHL	275	QY	1892	PSLPLSRTRRNRVSVYATWG-DHENRSPVKEPVEQPVTRKRLERLELOEAAAVPTPR	1950
975	AVDLEKLEARKRRFADSNLKAQKQKPEYKSSPEMED-ARVLSKKQDPDVSSREVILLREG	1033	Db	1172	SNESISEQFRTREODVEVLEPLKCELVSGESTGNCEDRLPVKGTGTEANGKKPSQOKLE-E	1230
276	ECVKPPL--EEPEDEMOCEVCVAHKVPGVTDVVAEIQKKNKPYIRHPEIGYDRSRKYWF	333	QY	1951	GRPKTRRRADDEEENEAKPEAE-TLKPPEGWSPRSQKTAAGGQPGQKGNKPEKVDAT	2009
1034	BAERPVRKEILLKRESKKIKLDRLNTVASPKDC-QELAS-----ISVSGSRPSSD	1083	Db	1231	RP--VNKSDQIKLKNTT-----DKNNENRESEKKGQRTSFOT-----	1268
334	LNRRL--IIEEDTENENEKKIWIYSTKVQIAELIDCLDKDYWEAEJCKIILEM---REEI	388	QY	2010	RPEATTEVGPOIGVKESMEPKAAEEAGSEQQRDRKDACTDKNPETAPVVEVKKPAP	2069
1084	LOARLGEIAGESVQEQSKPIPSKQLKQL-----QVLDQOGEREDV	1129	Db	1269	-----NGDKNPKIYL-KGECLKEISES---RVVSGNVPEPKVNNINKIIPENDIKSUT	1317
389	HR-HMDITEDLTNKARGSNKSFLLAANEIILSTRAKKG-DIDNVKS---PETEKDKNE	443	QY	2070	EKNSKXKRGSRNRLAVDKSASLKNVDAAVSPRGAAAQAAGERESGVAVSPEK---SESP	2127
1130	RKNYCSLRDETTERKSGKEGSHSVNTEKI-----GTDIDHTQSYRKQMEQSRKQ	1181	Db	1318	VKESAIRPFGINDVI-----MEDFNERNSETKSHL-----LSSSDAE	1355
444	TENDSKDAB-----KNREEFEDOSL-----EKSDDKTIPDDDPQKG-SEVGDF	486	QY	2128	QKEDGLSQLKSDPVDPEKEDEVSASGPSPEATQIAKQMELOEAOVHIAKLAASAS	2187
1182	MEMETAKEKFGKPKQDVDEYERRSLVHEVGKPPQDVTDOSPSPSKKRMHDVDFICTKR	1241	Db	1356	GNRY-DSLETL-PSTKESDSTQTTTPSASCPSNSVQV-----	1392
487	-KSEKSNGLSE-----SPCAGKA-----SGSTRITILRLNP--DSKLSQLKSOQ	529	QY	2188	RAYKADAPLEGLAPEDROKPAHOASETELAAAGSIINDISGEPEFPAPPYPGESOTDL	2247
1242	ERNYRSROISEDSRBTQSVVRHSGPHEDEDPGSPRLLSVKVGSPKVDKVLPSYNT	1301	Db	1393	-----EDME-----IETSEVKVTSPTSEESNLNDFIDENCLPIN	1431
530	VAAAHAENKLFKEGKVL-----VNASGEISRLST-----K	562	QY	2248	QPPAGAAQALOPSEEGMETDEAVSGILETEAATESRPPVNPADPSPAGPDTKREANSSE	2307
1302	VREESLKENPYDSSRREQMADMAKIKLSVLNSDELNRWDSOMKODAGREFDVSPNSIIK	1361	Db	1432	KNENVNGESKKKTIVETVMTSTVA-----TESKTVIVKEGDKQTVTVSSSTE	1479
563	KEVIMKGNINYY-----FKIQEOKYRYVINOYSTNSFALNKH-----QHRED	605	QY	2308	TSHSVPEAKGSKEV--EVTLVKDKGQKTRSRKRNKTKVAVPVESHVPSNOAQGE	2365
1362	RDSLRKRSVRDLEGPVPSDSEDEGE-----HKSHSPRASALYESRSLFLRDRDCKLRE	1417	Db	1480	NCAKSTVTTT-----TTTVTKLSTESTGSGVDIVISVKQSKT	1516
606	HDKR-----RLHAKHFCU-----TPAGEFKWNGSVHGSVKLTIITLRLTITQLENNIPSS	655	QY	2366	SPAANEGTVQHPAEQOEKQSEKPHSTPQSCSTDLSKI--ESTENSQOEISVEERTPT	2423
1418	RDERLSSSLERNKFYSFALDKTIP-----DTKALLERAKSLSSREEN---WS	1463	Db	1517	VVTVTVDLSLTTGTGLTVMSTVSKYSTDKVKVLMKFSRPPKTRSGTALPSYRKFVTKS	1576
656	FLHPNWAHRANIKAVQCMSPREFALAILAIECAKVPVVMPLPIWREFLGHTRLHRMTS	715	QY	2424	--KASVPDLP--POPAPVDEEPOARFVRHSHIIESDPVTPPSDPSIP-----	2468
1464	FL--DWDSRFANFN-----NKDKE-----KVDASAPRI--PSW-YMKKKKIR---	1501	Db	1577	TKKSIIVLPNDLKKLA---RKGGIREVYFNYNKAPALDIWIYPPSPRPTFGTWRYLQ	1633
716	IEREEKVKKKEKQEBEETMQATWVKYTFPVKHQVWKQGEYRVTYGGGWWSIKT	775	QY	2469	-----IPTLPSVTAAKLSPPVASGGI-----PHQSP-PTKVTEWITROE	2506
1502	TDSEGMDDKEDHEEGERGE-----WDISKITEK---KDOOVKELLDSDS-----DKPC	872	QY	1634	TVKSLAGVSLMLRLLWASLRWDDMAAKVPPGGGSTR-----TETSETEITTEIILKRRD	1687
776	HVYRFVPKLPNGNTNYNRKSLEGTKNNDENMDSDKRCRSRPPKTKIEPDSE--KDEV	833	QY	2507	EPRAOSTPS-----PALPDPTRKASDVDTSSSTLRKILMDPKYVTSATSVTSSV----	2554
1526	FASRF-----LHSSIFEQDSKRLQHLERK---EEDSDFIGRI	1560	Db	1688	VGPIYIRFEYCIRKIIICPIGPETPKETPTPQRKGLRSSALRKPRTPKQTGPVFIETW	1747
834	KGDAAKGADONE-----WDISKITEK---KDOOVKELLDSDS-----DKPC	872	QY	2555	-----TTALAEPVSAAPCLHEAPP--PVDSSKKPLEEKTAPPV-----	2590
1561	YKQTSSEGANSTTDSIQEPVILFHSRFEMLTMOQKEKQDQKPEVEKQEDTENHPKTP	1620	Db	1748	VAEELELWEIRAFARVEKEKAQAV-----EQAKKLEQQKPTVI	1789
873	KEEPMVEDDMKTESHVNCQESSQDVVNVSGEFLHRTSYKKTKSKSLDGLL---BRR	929	QY	2591	TNNSEIQASEVLVAA---DKEKVPVIAPIKITSVISRMPVSIDLENSQKITLAKPAPQTL	2647
1621	ESAPENKDSELTPPSVG---PPSVTVVLTESA---PSALEKTTGDKTVEAPLVEEKT	1674	Db	1790	ATSTTSPSTSTTSTISPAQKVMAVPIGSGVTGKVKVLTTKVGSPAPVTFPQKNFHOIF	1849
930	QKPTLEEK-----QRLEKIKLEGGIKGKTSTNSKNLSESPVITAKEGCQ	977	QY	2648	TGLVSAITGLVNVSLVPV-NALKGPVKGSVTT-LKSLVSTPAG-PVNVL-----	2693
1675	EPATVSEEAKPASEAPAPAVEQLEQVDLPPG-----ADPDKEAAMMP--AGVEEGSS	1724	Db	1850	ATWVKQGOSEN---SGVYQV---QOKVLGILPSSGTGTSQOFTTSFQPTATVTRPN--TS	1901
978	-SDSM---ROEQSPNANDQPEDLLQGCQSQSDSSVLRMSDPSTHTNKLKYP---DRVLD	1029	QY	2694	-----KGPVNVLTGPVNVLTTPVNAVTVGTVNAAPCTVNAASAVNATASATVTVAGAVTA	2748
1725	GDQPPYLDKPTTPGASFSQAESNVDPDPDSTQPIUSKPAQKSEANE--PRAEKPDATAD	1782	Db	1902	GSGGTTNSQVIT-----GPOIRPGM-----	1922
1030	---DVSIR---SPETKCPKQNSIENDIEEKVSDLASRGOEPTKSK-----	1068	QY	2749	ASGGVTATGTGVTVMAGAVIAPSTCKQKQASANENSRFHPGSMFVIDRPPADAGSAGLRV	2808
1783	AEPDANKAEAPESOPASEDELYD-----PPVAAKDKKPKNKSRSKTPVQAAAASV					

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 01:17:18 : Search time 105.119 Seconds
(without alignments)
3939.095 Million cell updates/sec

Title:
Perfect score: 14333
Sequence: 1 MYSEEEEDGDAETQDSE.....KLKGFKASRSHNNKLQSTAS 2781

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 556269 seqs, 14893369 residues

Total number of hits satisfying chosen parameters: 556269

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Result No.	Query Match	Score	Length	DB ID	Description
1	1474	10.3	328	9	US-09-925-297-816
2	982.5	6.9	238	9	US-09-729-835-80
3	430.5	3.0	3664	15	US-10-177-293-423
4	418	2.9	4019	10	US-09-738-973-425
5	418	2.9	4019	10	US-09-854-133-425
6	418	2.9	4019	15	US-10-144-649A-425
7	416.5	2.9	2665	9	US-09-864-761-34248
8	403	2.8	3256	10	US-09-919-172-98
9	403	2.8	3256	11	US-09-919-039-21
10	401.5	2.8	1367	10	US-09-801-368-108
11	401	2.8	2344	9	US-09-815-242-12713
12	391	2.7	5179	9	US-09-922-217-1068
13	391	2.7	5179	10	US-09-833-263-1068
14	391	2.7	5179	14	US-10-025-380-1068
15	384.5	2.7	3899	15	US-10-171-311-4

Sequence 8, Appli

Sequence 2, Appli

Sequence 6, Appli

Sequence 20, Appli

Sequence 83, Appli

Sequence 5816, Ap

Sequence 12967, A

Sequence 8, Appli

Sequence 11, Appl

Sequence 109, App

Sequence 20, Appl

Sequence 1, Appli

Sequence 32, Appli

Sequence 30, Appli

Sequence 2, Appli

Sequence 7, Appli

Sequence 4, Appli

Sequence 2, Appli

Sequence 7, Appli

Sequence 2, Appli

Sequence 3, Appli

Sequence 52, Appli

Sequence 8, Appli

Sequence 299, App

Sequence 149, App

Sequence 82, Appli

Sequence 8, Appli

Sequence 2, Appli

Sequence 6, Appli

Sequence 20, Appli

Sequence 83, Appli

Sequence 5816, Ap

Sequence 12967, A

Sequence 8, Appli

Sequence 11, Appl

Sequence 109, App

Sequence 20, Appl

Sequence 1, Appli

Sequence 32, Appli

Sequence 30, Appli

Sequence 2, Appli

Sequence 7, Appli

Sequence 4, Appli

Sequence 2, Appli

Sequence 7, Appli

Sequence 2, Appli

Sequence 3, Appli

Sequence 52, Appli

Sequence 8, Appli

Sequence 299, App

Sequence 149, App

Sequence 82, Appli

ALIGNMENTS

RESULT 1

US-09-925-297-816

; Sequence 816, Application US/09925297

; Patent No. US20020081659A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PAl05

; CURRENT APPLICATION NUMBER: US/09/925,297

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05989

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 928

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 816

; LENGTH: 328

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (170)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (172)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (174)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (178)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (183)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (269)

THIS PAGE BLANK (USPTO)

Db 1120 SATSA-----SAGGTATPGKPPA--VVSQAAGS--TTVCATLTSVSTTSPSTA 1168
QY 1547 DKVKLMKFSRPKKTSRGFTALPSYKFTVKTKSKSIFVLPNDLKKLARKGGIREVPYFNY 1606
Db 1169 SOLSI-----QLSSSTSTPLAETVVVSAH----- 1193
QY 1607 NAKPALDIWPPSPRPTFGIYWRYLQTVKSL--AGVSLMLRLWLASLRWDDMAAKVPPG 1664
Db 1194 ----SLDKTSHSS--TTGLAFSLAPSSSSSPGAGVSSYIS-----QPG 1231
QY 1665 GGSTRTFSEYETITTEILKRRDVGPIGIREYCIKIIICPIGVPEPKETP--TPORKGL 1723
Db 1232 G-----LHPLVIPSVIASPTILPQAAGP 1254
QY 1724 RSSALRPKRPEPKOTGPV-----IETWVAEE-----ELE----- 1754
Db 1255 TSTPLLPQVPSIPPLVQPVANVPAVOQTLIHSQPQALLPNQPHTHCPEVDSDTQPKAPG 1314
QY 1755 LWEIRAFARVE---KEKAQAVEQAKRLEQ---KPTV---IATSTTSP---SSTTS 1802
Db 1315 IDDIKTLSEKLSRFSHSSGAQHASVLSLETSLVISTVTGPIPTTAVAPSKLLTSTS 1374
QY 1803 TISPAQKV---WVAPIGSVTTGKMWLTTKVGSPPATVTFQONKNFHQTFATVVKOGQS 1858
Db 1375 TCLPPTNLPLGIVALPVPVTPG-----QVSTPVSTT----- 1407
QY 1859 NSGVVOVQOKVLGIIPSSGTSGQTTFSQPTATVTRPNTISGGGTTNSQVITGPQI 1918
Db 1408 -----TSGVKPGTAPSKPPLTKAPVLPVGTGTELPAGTLPSEQLPPFP-- 1448
QY 1919 RGMVTVIRTPLOQ-----STLGRAIIRTPVMVOPGA--POOVMTQIIR 1959
Db 1449 GPSLTOSQOQPLEDDAQLRRTLSPEXITVTSAGVPVSMAPTAITEAGTQPKGVSOYKE 1508
QY 1960 GPVSTVIAVSA-----PNTVSSSTPGQKS-----LTSATSTSNIOSSASQPP 1999
Db 1509 GPVLATSSGAGVFKMGRFQVSAADGAQKGNKSEDAKSVHFESSTSESSVLSSSS--- 1565
QY 2000 RPOQGVKLTMAQLTOLTQGHGNGQLTVVIOGOGOTTQOLQILPOGTVLPCPGQOLMQ 2059
Db 1566 -PESTLVKPE-----PNGITI--PGISSDVPE 1589
QY 2060 AAMPNGTVQRFLEPTLATATATTT-----TTVSTTAAGTGEQRQSK----- 2103
Db 1590 SAH-----KTTASEAKSDTGQTKVGRFQVTTANKVGRFSVSKTEDKITDT 1636
QY 2104 -----LSPQMQVHQDKTLPPAQSSSVGPAKAQPTAQPSARPQOTQPO----- 2147
Db 1637 KKEGVPVAPPFMDLEQAVLP-----AVIPKKEKPELSEPHLNGPSSDPEAAFLSRDVED 1691
QY 2148 ---SPAQP-----PG----- 2152
Db 1692 GSGSPHSPHOLSKSLPSONLSLNSFNSSNYSNDSNEDIEDLKLRLRLRDKHLK 1751
QY 2153 EVQ-----TOPEVQOTOTVSSHVP-----SEAQPHASQSKPVAA---QSQ 2191
Db 1752 EIQDLQSRQKHEIESLYTLKLGKVPVAVIIPPAAPLSGRRRRPTKSKGSKSSRSSLGNKS 1811
QY 2192 PQ--SNVQGSPPRVQSPQSTRIRPSTPQLSPGQSOVQTTTSQPIPIQHTSLOIPSQ 2249
Db 1812 POLSGNLGGQSAASVHLHPQOT--LHP-----PG-----NIPES 1842
QY 2250 GPOQ--SOPQVS--STQTLSSGQTLN--QVSVSSPSRPOLOIQOP 2289
Db 1843 GQNLQLPLKPSFSSDNLISAFSTDGALSVPSLSAPCGGIKQP 1885


```
Db 971 -----AQPV-----HDLVSVILASPTTEQPTATVTIADSGGDVQP- 1006
Qy 1816 SGSVT-----TGTWMLTKV-----GSPATVTFQKNKFNHQTATWVKQ--G 1856
Db 1007 -GTVTLVCSNPPCETHETGTTNTATTVVANLGGHPQPTQVQFVCDRQEAASLVTSTVG 1065
Qy 1857 QNSGVVQVQKVLGIIPSSGTSTGTSQQTFTSQPTATVTIRPNTSGSGGTTNSQVITGP 1916
Db 1066 QQSGSVVRVCSNP-----PCET-----HETGTTNTATTAT-----SNMAGQHCNS- 1106
Qy 1917 QIRPGMTVIRTPLOQSTLGKAIIRTPVMVQVQAPQ--VMTQIIRGQPVSTAVS-APNTV 1973
Db 1107 -----PCETHETGTTNTATTAMSSVGNHQRDARRACAAAGTPAVIRISVATGAL 1156
Qy 1974 SSTPGQKS-----LTSATSTSN-----QSSASQPP-----RPOQG 2004
Db 1157 EAAQGSKSQCOTROTSATSTMTVMATGAPCSAGPGLLGPMSAREPGGRSFAVLQAPLSS 1216
Qy 2005 QVKLTMAQLTQTOGHGNGQLTVVIOGQTTQQLQIP-----QG-----VTVLP 2051
Db 1217 KVRLLSPSIKDLPAGRHSHAVSTAAMTRSSVSGAGEPRMAPVCESLOGGSPSTTVTTVALE 1276
Qy 2052 G---PGQQLMQ-----AAMPN-GTVQRFLETPLATATATASTTTTTS 2090
Db 1277 ALLCPSTVTVQCSNPPCETHETGTTNTATTSNAGSAQRVCSNPPCETHETGTTHTATTA 1336
Qy 2091 TTAAGTGEQRQSKLSP---QMQRHDKTLPPAQSSVG-----PAKAQPT-----AQ 2135
Db 1337 TSNGGTGQEGGQPPAGRPCETHQTTSTGTTMSVSVGALLPDATSSHRTVESGLEVAAA 1396
Qy 2136 PSARPQPTQPOSP-----AQEVOTQP--EVOTQTTVSHVPSVSAQPTHAQSSKPQV 2186
Db 1397 PSVTPQAGTALLAPFTPTQRVCSNPPCETHETGTTHTATTTVTSNNSSNQDPPPAASDQGE 1456
Qy 2187 AAQSQPOSNOGQSPVRVQSPQ-----TRRPSTP---SOLSPGQSOVQTTTSOPIPI 2238
Db 1457 ESTQGSVNITSSAIIITVSSITLRAVTVTQSTPVPGRSVPPPELQVSGPRQQLP- 1515
Qy 2239 QPHTSLQIPGQPOQPOVQSSQTLSSGQTLNOVSVSPSRPQ-----LQIQQ 2288
Db 1516 -PRLQASASTALMGESAELVLSQSOTPELPAAVDLSTGTPSSQESAGSAAVAVTVVQV 1574
Qy 2289 POPQVIAVPOLOQOVVLQIQSOVV-----AQIQAOQSGVPOQIK 2329
Db 1575 PPTQSEVDQLSLPQELMAEAQAGTTLMVTGLTPEELAVTAAEAQAAATBEAQALA 1634
Qy 2330 LQLPIQIQQSAVOTHQIQNVTVVQAASVQELQVQOLRDQOQKKQOQIEIKREHTLQ 2389
Db 1635 IQAVLQAAQAVMGCTGEPMDTSEAAAVTQALGHL-SAEQEQCAATTIPVLTQOEELAA 1693
Qy 2390 ASNOSEIIQOVVMKHNVAIEHLKQKKSMTPAE 2422
Db 1694 LVQOQQLQEAQAQOQH-----HHLPTALAPAD 1721
```

RESULT 13

```
US-08-393-703-5
; Sequence 5, Application US/08393703
; Patent No. 5585239
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; NUMBER OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
```

```
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,703
FILING DATE: 24-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-2/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-393-703-5
```

```
Query Match 2.5%, Score 353; DB 1; Length 2035;
Best Local Similarity 21.4%, Pred. No. 2,le-12;
Matches 264; Conservative 146; Mismatches 515; Indels 308; Gaps 48;

Qy 1368 TKESDSTQTTT-PSACPCESNSNQVEDMEIETSEVKKVT-----SSPITSEESNLSN 1420
Db 619 TSVSSATNTSRPIITVHKSGTVTAQQAQVTVVVGCVTKTITLVKSPISVPVGSAL-- 676
Qy 1421 DFIDENGLPIKNENNVNGESKRKTVITTEVTMTSTVATESKT-----VIKVE---KDKK 1471
Db 677 -----ISNLGKVMVSVQTKPVQTSVAGTQAGSTGPTVQIIQKGPLPAGTI 721
Qy 1472 QTVVSSSTENCAKSTVTTTITTTVTKLSTPSTGSGVDIISKBQSKTVTTVTYDTS----- 1525
Db 722 LKLVTSAD--GKPTTIITTTQASGAGTKPT--ILGISSVSPSTTKPGTTIIKIIPMSAI 777
Qy 1526 LTTTGGTLVTSMTVSKYSTRDVKVLMKFSRPPKTRSGTALPSY-----RKFTVSKTKKS 1580
Db 778 ITOAGATGVTSSPGIKSPITITITTKVMTSGTGAPAKIITAVPKIATGCGQGVTVVVLKG 837
Qy 1581 IFVLPNDDLKLRKGGIREVPYFNNA-KPALDIWPYPSPRPTFGITWRYRLQTVKSLA 1639
Db 838 APGQPGTILRTVP-MGGVRLVTPVTVSAVKPAVTVTLV---KGTGVT---TLGTVTGTV 890
Qy 1640 GVSMLRLWLASLRWDDMAAKVPPGGGSTRPTSETETITTEIIRKRDVGPYIGRFEYCI 1699
Db 891 STSL-----ACAGGHSTASLATPITTLGTIATLS----- 920
Qy 1700 RKIICPIGVETPKETPTPQRKGLRSSALRPKRETPKQ-----TGPVITETWAEEL 1755
Db 921 SOVINPTAITVSAAOITLTAAAGLTTTITTMQPVSOPTQVTLITAPSGVE----- 970
Qy 1756 WEIRAFARVEKEKAQAEQAKKRLQEQKPTVIATSTTSPTSSTTSIPSPAOKVMVAPI 1815
Db 971 -----AQPV-----HDLVSVILASPTTEQPTATVTIADSGGDVQP- 1006
Qy 1816 SGSVT-----TGTWMLTKV-----GSPATVTFQKNKFNHQTATWVKQ--G 1856
Db 1007 -GTVTLVCSNPPCETHETGTTNTATTVVANLGGHPQPTQVQFVCDRQEAASLVTSTVG 1065
Qy 1857 QNSGVVQVQKVLGIIPSSGTSTGTSQQTFTSQPTATVTIRPNTSGSGGTTNSQVITGP 1916
Db 1066 QQSGSVVRVCSNP-----PCET-----HETGTTNTATTAT-----SNMAGQHCNS- 1106
Qy 1917 QIRPGMTVIRTPLOQSTLGKAIIRTPVMVQVQAPQ--VMTQIIRGQPVSTAVS-APNTV 1973
```

```
QY 1487 TTTTITVTKLSTPSTGSGVDIISVKEQSKTVVTTTITVTDLSLTTGGTLVMTSVKEYSTR 1546
Db 1180 SATSA-----SAGSTATPGPKPPA--VVSQQAAGS--TTVGATLTSVSTTTSEPPSTA 1228
QY 1547 DKVKLMKFSRPKKTSGTALPSYKRVKTKSKSIFVLPENDDLLKLARKGGIREVPFNY 1606
Db 1229 SQLSI-----QLSSSTSTPTLAETVVSAAH-----1253
QY 1607 NAKPALDIPYSPRPRTFGITWRYRLQTVKSL--AGVSLMLRLLWASLRWDDMAAKVPPG 1664
Db 1254 ----SLDKTSHS---TTGLAFSLSPSSSSSPGAGVSYIS-----QPG 1291
QY 1665 GGSTRFETSETITTEILIKRRDVGYPYIRFEVCIRKIICPGVPETPKETP--TPQRKGL 1723
Db 1292 G-----LHPLVIPSIASTPLPQAAQP 1314
QY 1724 RSSALRPKRPETPKQGPV-----IETWVAEE-----ELE-----1754
Db 1315 TSTPLLPQVPSIPPLVOPVANVPAVQOTLIHSQPQALLPNQPHTHCPEVDSDTQPKAPG 1374
QY 1755 LWEIRAFARVE---KEKAQAVEQAKKRLQEQ---KPIV---IATSTTSPT---SSTWS 1802
Db 1375 IDIKTLEEKRLSLFSEHSSGAQHASVLSLETSLVIESVTGPIPTAVAPSKLLTSTS 1434
QY 1803 TISPAQKV----MVAPISGSVTTGTRKMLTTKVGSPIATVTFQONKFNHQTATWVKQGS 1858
Db 1435 TCIPTNLPLGTVALPVPVTPG-----QVSTPVSIT-----1467
QY 1859 NSGVVQVQKVLGIIPSSGTGSOQTFTSFOPRATVTRPNTSGSGGTSTNSOVITGPQI 1918
Db 1468 -----TSGVKGTAPSKPPLTKAPVLPVGTLPAGTLPSEQLPFPF--1508
QY 1919 RGMVTVRTLQO-----STLGKALIRTPVMVPGA--PQOVMTQIIR 1959
Db 1509 GPSLTOSQQLDLDLQRLTSLPEXITVTSAVGPVSMAPTAITACTOPQKGVQVKE 1568
QY 1960 GQPVSTAVSA-----PNTVSTSPQKS-----LTSATSTSNIOSSASQPP 1999
Db 1569 GPVLATSSGAVFKMGFRQVSVAAQGAQKEGKNSEDAKSVHFESSTESSVLSSSS---1625
QY 2000 RPOGQVKLTMAQLTQLTOGHGNGQLTVVIOGQGTGQLQLIPQGVTVLPQGOQLMQ 2059
Db 1626 -PESTLVKPE-----AVIPKKEKPELSEPHLNGPSSDPEAAFLSRDVED 1751
QY 2104 -----LSPQMQRHDKTLPPAOSSSVGPAAQPTAQAQSPQPTQPO-----2147
Db 1697 KKEGPVAPPPFMDLEQAVLP-----AVIPKKEKPELSEPHLNGPSSDPEAAFLSRDVED 1751
QY 2148 ---SPAOP-----2152
Db 1752 GSGSPHSPHOLSSKSLPSQNLSSLSNFSNYSNMSNEDIEDDLKLELRRLDRKHLK 1811
QY 2153 EVQ-----TOPEVQOTQTVSSHVP-----SEAQPTHAQSSKQPVAA---QSQ 2191
Db 1812 EIQDLQSRQKHEIESLYTKLGKVPYPAVLIIPPAPLSGRRRRPKSKGSSRSSLGNKS 1871
QY 2192 PQ---SNVQGSPPVRQSPSTRIRPSTPSQLSPGQSQVQVTTSTQPTIPQHT-----2242
Db 1872 POLSGLNLGGQSAASVLHPQOT--LHP--PGNIPESGNQL-----LQPLKPSPSDNLXSAF 1924
QY 2243 ----SLQIPSGQPOQPOVQSSQTOTLSLSSGQTLNQVSVSPSRPOLQIQPOP 2291
Db 1925 TSDGAISVPSLSAP---CQGTSTSTNTV--GATVNS-----QAAQAP 1961

Query Match 2.5%; Score 353; DB 1; Length 2035;
Best Local Similarity 21.4%; Pred No. 21e-12;
Matches 264; Conservative 146; Mismatches 515; Indels 308; Gaps 48;

QY 1368 TKESDSTQTTT-PSACSPESNVQVEDMEIETSEVKKVT-----SSPTSEESMLSN 1420
Db 619 TSVSSATNTSTRPIITVHKSGTVTAQQAQVWTVVGVGVTKITLTKSPISVPGGSAL-- 676
QY 1421 DFIDENGLPINKNENGVESKRTVITEVITMTSTVATESKT-----VIKVE---KGDK 1471
Db 677 -----ISNLGKVMVSVQTKPVQTSVAVTQQAQSTGPTVQIITKGPLPAGTI 721
QY 1472 QTVVSSPENCASKSTVTTTTTTTNTKLSPTSGGSDIISVKEQSKTVVTTTVDTS-----1525
Db 722 LKLVTSD--CKPTIITITQASGAGTKPT--ILGISSVSPSTTKPGCTTIITKIPMSAI 777
QY 1526 LTTTGGTLVMTSVKSEYSTROKVKLMKFRPKKTRSGTALPSY-----RKFTVKSTKS 1580
Db 778 ITQAGATGVTSSPGIKSPIITITKVTMTSGTGAPAKIITAVPKIATGHGQGVTVVYLGK 837
QY 1581 IFVLPNDDLLKARKGIGIREVPFNYNA--KPALEDIPYSPRPRTFGITWRYRLQTVKSLA 1639
Db 838 APGQPGTILRTPV--MGVRLVTPVTVSAVAPVTLVW---KGTGTV---TLGTVTGTV 890
QY 1640 GVSIMLRLLWASLRWDDMAAKVPPGGGSTRTETSETTEITTEIKRRDVGYPYIRFEYCI 1699
Db 891 STSL-----ACAGHSTASLAPITITLGIATLS-----920
QY 1700 RKIICPIGVPTPKETPTPQKRLRSSALRPKRPETPKQ---TGPVIETWVAEELEL 1755
Db 921 SQVINPTAITVSAOQTLTAAGGLTTPITMQPVSQPTQVTLITAPSGVE-----970
QY 1756 WEIRAFARVEKEKAQAVEQAKKRLQEQKQKPTVATSTTSPTSTTSTISPAQKVMVAPI 1815

GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Willson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/046,585
; FILING DATE: 12-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57503-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-046-585-5
```

RESULT 12

US-08-046-585-5

; Sequence 5, Application US/08046585

; Patent No. 5453362

Qy	2000	RPQGGVKLTWAOI	TQLTQGHGNGQLTVV	YIQGGOTTGOLQ	LIPGCVTLPGPGQOLMQ	2059
		:-: :-:				
Db	1566	-PESTLVKPE-				
		:-: :-:				
Qy	2060	AAMPNGVQREFLT	PLATTATTASTTT-			2103
		:-: :-:				
Db	1590	SAH-----	KTASEAKSDGQTKVGR	QCVTTANKVGRFVS	YSKTEDKITDT	1636
		:-: :-:				
Qy	2104	-----	LSPOMQVHQDKTL	PPAOSSSVGP	AKAQPQAQPSARPQ	2147
		:-: :-:				
Db	1637	KKEGPVASPPF	MDLEQAVLP--			1691
		:-: :-:				
Qy	2148	---SPAQP----				2152
		:-: :-:				
Db	1692	GSGSPHS	PHOLSKSLPSQ	LSQSLNSFN	SYMSSDNESDIED	1751
		:-: :-:				
Qy	2153	EVQ-----	TOPEVQTOTV	SHVP-----		2191
		:-: :-:				
Db	1752	EIODLSQ	RKHETESLYTK	LGVPPAVI	IPPAAPLSGR	1811
		:-: :-:				
Qy	2192	PQ--SNVQGS	VRVQSPSOTR	IRPSTPSQL	SPQGSQVQTTTSQ	2242
		:-: :-:				
Db	1812	PQLSGNLS	GGSAASVLHP	QQT-LHP-PG	NIPESGQNL---	1864
		:-: :-:				
Qy	2243	----SLQIP	SGQGPQSGPQ	VSQTSOTL	SSCQTLNQVSVSS	2291
		:-: :-:				
Db	1865	TSQCATSV	PLSAP---G	CGTSTNTV	-GATYNS-----	1901
		:-: :-:				

RESULT 11

```

US-09-854-856-30
; Sequence 30, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 2108
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2108)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-30

```

[illegible]

Qy	576	KLQGEGRYRYHNOYQSTNSPALNKHQHRED-----HDKRRRLAHK-----FCLTPA 621
Db	302	ELMTSGTLKTYLKRFKVMKIKVLRSKMCQILKGLQFLHTRTPPIIHRDLKCDNIFITGPT 361
Qy	622	GEFKWGSVHGSKVLTISTLRLITQLENNIPSFHLPNWASHARWIKAV-----672
Db	362	GSVKIGD-----LGLATLK-----RASFAKSVIGTPEFMAP 392
Qy	673	QMSCKPRE-----FALALATLECA-----VKPV-----VMLPIW 701
Db	393	EMYEKDESDVYAFQCMLEMATSPYPYSECNAQAIIYRRVTSVKPSPAFDKVAIPEV 452
Qy	702	RE-----FLGTHRLHRMTSI-----EREKEK-----VKK 727
Db	453	KEILEGICRQNKDERYSIKDLLNHPAQEETGVRVELABEDDCEKIAIKLWLRIEDIKKL 512
Qy	728	EKQEBEETWQATWKYTFPVKHQWVKQGEYRYTGY---CGWSWISKTHYRVPKPL 784
Db	513	KGRYKDEA-----IEFSFLDERDPVEDVAQEMVSEGYVECDGHDHTMAKA-----557
Qy	785	PGNTNVNRYKSLEGTKNMNDENDESDKRCSRPKKIKITEPDEKDEKVGSDAAKGAQD 844
Db	558	-----IKDRVSLIKRKEQQLVREBQEK-----KQESLSLKQQVEQSASQ-----600
Qy	845	NEMDISKITEKKDQDVKELDSDSDRCKEPEPMEVDMDKTESHVNCQESSQVDVYVNSE 904
Db	601	-----TGIKQLPSASTGIP-----TASTTSASVSTQVEPEEPA 634
Qy	905	GFHLRTYKKTKSSKLGDGLLERRIKOFTLEEKORLEKIKLEGIGIKGTSTNSKNLS 964
Db	635	DOHQOLQYOQPSISVLSDGTVD-----SQGSSSVFT 665
Qy	965	ESPVITRAKEGCQSDMRQEQSPNANDOPEDLIOGCSQSDSVLRMSDPSHTNKLYPK 1024
Db	666	ESRV-----SSQQTVYSGSHEQAHTGTVPGHIPSTVQAOSQB-----HGVIYP 710
Qy	1025	DRVLDDYSIRSPETKCPQNSIENDI-----EEKVSDLASRQGEPT-----KSKT 1069
Db	711	SSVQOGI-----QQTAPQQTQVQYLSQTSSEATTAQPVSQPAQVLPQVSAACKQST 765
Qy	1070	KGNDFDIDSKLASADDIGPLICKNNKPLQEBESDTIVSSSKSALHSSVPKSTNDRDAP 1129
Db	766	QG-----VSQVAPAEVAV-----AQP---QATQPTTLASSVSDSAHSDVASGMSDGENV 812
Qy	1130	LSRAMDEPK-----LGCDESENSTLENSSDT-----VS 1158
Db	813	PSSSGRHEGHTKRRHYKKSVRSRHRBKTSRPLRILNVSKGRVVECOLETHNKKMWT 872
Qy	1159	IQ-----DSSEE-----DMIVQNSNESISEQFRTREQDVEVLPEKLCELYSGESTGN 1205
Db	873	FKFDLDGNEPEIATINVNDFILATERESFVDQVR---EIEKADEMLSEDSVVEPEGD 929
Qy	1206	CEDRLPVKGYTEANGKPKSOQKLEER---PVNKCSDQIKLKNTTDDKNNENRESEKKGQR 1262
Db	930	-QGLIESLQGGKDDYGFSGSQLEGEFKOPIPASSMPOQIGIPT-----970
Qy	1263	TSIFQINGKDKPKIYLKGE--CLKEISESRVYSGNVEPKVNNIKIIPENDIKSLTVKE 1320
Db	971	SSLTQV-----VHSAGRRFIVSPVPERLSRESKVFP-----SEITDVVAAS 1011
Qy	1321	SAIRPFINGDVIMEDFNERNSSETKSHLLSSDAEG-----NYRDSLETLPSTK---ES 1371
Db	1012	TAQSPGN---LHSASSLSLQAFSELRAQWTEGPNATPPNFSHTGTFTFPVVPFELS 1068
Qy	1372	DSTQTTTPPSACPESNVQVEDMEIETSEVKKVTSPITSEESNL-----SNDFIDEN 1426
Db	1069	IAGVPTTAAATAPVATPSSPND--ISTS---VIOSEVTVPTVEGIAGVATSTGVVTS 1122
Qy	1427	GLPINKNENYNGESKRKTVITEVTTMTSTVATESKTVIKVEKDGKQTVVSSSTENCAKSTV 1486
Db	1123	GLPT---PPVSESPVLSSVSSSITIPAAVSGISITSPSLQVPTSETVSSSTALYPSVT 1179

QY 622 GEFKNGSVHGSKVLITSTLRLTITOLENNIPSPSLHPNASHANWIKAV----- 672
Db 362 GSVKIGD-----LGLATLK-----RASFAKSVIGTFEPMAP 392
QY 673 QMCKSPRE-----PALALAILECA-----VKPV-----VMLPIW 701
Db 393 EMYEKYDESVDVAFGCMLEMATSEYPYSECNAQIYRRVTSGVKPPASFDKVAIPEV 452
QY 702 RE-----FLGHTRLHMTSI-----EREKEK-----VKKK 727
Db 453 KEIIIEGCIQONKDERYSIKDILNHAFOETGVRVELAEEDGKIAIKLWLRIEDIKKL 512
QY 728 EKKOEETMQOATWVYKTFPVKQVWKQKEEYRTGY-----GWSWISKTHVTRFVKPL 784
Db 513 KGKYKDNFA-----IEFSFLEKRDVPEDVAGVMEVSGYCEGDHKTMAKA----- 557
QY 785 PGNTNVNRYKSLGKTKNMNDENMDSEOKRKSRSPPKKIKIEPDSEKDEVKGSDAAGADQ 844
Db 558 -----INDRVSLIKRKRQOLVREEQEK-----KQESSLKQOQVESSASQ----- 600
QY 845 NEMDISKITEKKDQDVKELSDSDKPKCEPMEVDDDMKTESHVNCQESSQVDVNVNSE 904
Db 601 -----TGIIKQLPASASTGIP-----TASTTSASVSTQVPEPEPEA 634
QY 905 GFHLRTSYKKKTKSKLDGILLERIKQFTLEEKORLEKIKLEGGIKGIGKSTNSKNLS 964
Db 635 DOHQOLOQOQPSISVLSGTVD-----SGQSSVFTESRVSS 671
QY 965 ESPVITRAKEGQSD-----SMRQEQSPNANDQPEDLQGCQSQDSSVLRM 1011
Db 672 QQTV-----SYGQHEQAHSHTGVPHIPSTVQAOQSGHGVYPPSSVAAQOGQOGQSSSL 727
QY 1012 SDPHTTNKLYPKDRVLDVDSIRSPETKPKQNSIENDI-----BEKVSDLASRGOEP 1064
Db 728 TGVSSSOPIQHPOQ-----QGIIQTQAPPQQTVOYLSLQTSSTSEATTAAQPVSQPOAP 780
QY 1065 T-----KSKTKGNDFFIDDSKLASADDIGTILCKNKKPLIQEESDTIVSSKSLAHS 1116
Db 781 QVLPOVSAGKOSTG-----VSQVAPAEVAV-----AQD-----QATQPTLIASVDSAH 827
QY 1117 SVPKSTNDRADTPLSRAMDFEGK-----LGCDSESNSTL 1150
Db 828 DVASGMSDGNENVPSSGRHGRITKRRHYKSVRSRSHRHKTSRPLRLINVSKGRDV 887
QY 1151 ENSSDT-----VSIO-----DSSEE-----DMIVONSNESTSEFRTREQDVEVLEP 1192
Db 888 ECQLETHNRKMVTEKFDLDGDNPEEATIMVNDFFILAIERESFVDQVR-----EIIERKADE 944
QY 1193 LKCELVSGESTGNCEDRLPVKGTANGKKPQOQKLEER-----PVNKCSDQIKLANTTDDK 1249
Db 945 MLSEDDVSEPEGD-----QGLSEJOGKODYGFGSQKLEGEFKQIPASSMPQQIGIPT----- 998
QY 1250 NENRESEKKGQRTSTFQINGKDNKPKIYLKGE-----CLKEISESRVSGNVBPKNVNNINKI 1307
Db 999 -----SSLTQV-----VHSAGRFRFVSPDESRLRESKVEP----- 1029
QY 1308 IPENDIKSLTVKESAIRPFINGVIMEDFNERNSETKSHLLSSDAEG-----NYRDS 1361
Db 1030 ---SEITDTVAASTAQSPGMN-----LSHSASSLSLQQAFLSELRLAQMTGPNATPNFSHT 1083
QY 1362 LETLPSYK-----ESDSTQTTTPSASCPSNSVNOVEDMEIETSEKVKYTSSTITSEESNL 1418
Db 1084 GPTFPVVPFPLUSSITAGVPTAAATAPVATSSPPND-----ISTS-----VIQSEVTVPTBEGI 1137
QY 1419 -----SNDFIDENGLPINKNENVSKRKTIVITEVTTMTSTVATESKTIVIKVEGDKQT 1473
Db 1138 AGVATSGVTVSGGLPI-----PPVSESPVLSVSVSSIIPAVVSIITSPSLQOVTSISEI 1194
QY 1474 VVSSTENCAKSTVTTTTTTTKLSTPSTGGSDVLIISVKEQSKTVTTTTTTVTDLSLTGTGTL 1533
Db 1195 VVSSTALYPSVTVSATS-----SAGGSTATPGPKPPA-----VVSQQAAGS-----TTVGATL 1243
QY 1534 VTSMTVSKEYSTRDKVXKLMKFSRPKTKRSGTALPSYRKFTVTKTKKSIFVLPNDDLLKLA 1593

Db 1244 TSVSTTTTSPETASOLSI-----QUSSTSTPTLAETVVVSAH----- 1281
QY 1594 RKGIREVPYENYNAKPAIDWPPSPRPTFGITWRYRLQTVKSL-----AGVSLMLRLWLAS 1651
Db 1282 -----SLDKTSHSS-----TTGLAFSLSPSSSSPGAGVSYIS----- 1316
QY 1652 LRWDDMAKVPDPGGSTRTTSETTEITTIKRDVGPYGIREFYCIKRIKICIPGVPET 1711
Db 1317 -----QPGG-----LHPLVIPS 1329
QY 1712 PKETP-TPORKGLBSSALRPKRPETPKOTGPV-----LIETWVAEE----- 1751
Db 1330 IASTPILPQAAAGPTSTPLLPQVPSIPPLVQPVANVPVAVQOFTLIHSQPOPALLPNOPHTHC 1389
QY 1752 -----LWEIRAFARVE-----KEKAQAVEQOAKRLEQO-----KPTV-----IATS 1792
Db 1390 PEVSDTQPKAPGIDDIKTLEKRLSLFSEHSSGCAHASVLETSLVIESTVTPGIPPT 1449
QY 1793 TTSPT-----SSTTSTISPAKV-----MVAPISGVSTTKMVLTKVGPSPATVTTQONKNF 1845
Db 1450 AVAPSKLLTSTSTCLPPTNLPLGTVALPVTVPVTPG-----QVSTPVSTT----- 1495
QY 1846 HQTATWVKOGQSGVGVVQOVKVLGIIPSTSTGTSQOFTTSFOPRTATVTRPNTSGSGG 1905
Db 1496 -----TSGVKPGTAPSKPPLTKAPVLVPVGTLPAG 1525
QY 1906 TTSNSQVITGPQIRPGMTVIRTPLOQ-----STLGKAIIRTPVMVQPG 1948
Db 1526 TLPSEQLPFPF--GPSLTQSOQPLEDLDAQLRRTLSPXKITVTSAVGPVSMAPTAITEA 1583
QY 1949 A--PQVMTQIIRQPVSTAVSA-----PNTVSVSTPGQKS-----LTSAT 1986
Db 1584 GTQPKGVSVQKEGVPVLATSSGAGVFKMGRQVSVAAADGAQKGNKSEDAKSVHFEST 1643
QY 1987 STSNIOSSASQPPRPOQGVKLTMAQLTQLTQGHGNOGLTVIIOGOGOTTQOLQIPQG 2046
Db 1644 SESSVLSSSS-----PESTLVKPE-----PNG 1665
QY 2047 VTVLPGGQOOLQMAAMPNGTVQREFLPTLATTTASTTT-----TTVSTTAAGTGE 2098
Db 1666 ITI-PGISSDVPESA-----KITASEAKSDTGQPTKVGREFQVTTTANKVGR 1711
QY 2099 QROSK-----LSPQMVHODKTLPLPAQSSSVGPAKAQPTAQPSARQPQT 2144
Db 1712 FVSKTEDKITDTRKEGVASPPFMDLEQAVLP-----AVIPKKEKPELSEPHLNGPSS 1766
QY 2145 QPO-----SPAQP----- 2152
Db 1767 DPEAFLSRVDGSGSPHSPHQSLSPSONLSQSLNSFNSSNSMSSDNESDIEDDL 1826
QY 2153 -----EVQ-----TQPEVQTQTTVSSHVP-----SEAQPTHAQS 2181
Db 1827 KLELRLRDKHLKEIQDLSQKQHEIESLYTKLGKVPVPAIIPPAAPLSGRRRRPTKSKG 1886
QY 2182 SKPOVAA-----OSQPO-----SNVQOGQPVVRVQSQRTFIRPSTPSQLSPGQOQSVQTTTSQPI 2236
Db 1887 SKSRSSRLGNKSPQLSGNLNLSQSAASVYLPQQT-LHP-----PG----- 1925
QY 2237 PIQHTSLQIPSOQPO-SOPOVOS-SPOTLLSGOTLN-OVSVSSPSRQLQIOQP 2289
Db 1926 -----NIPESQONLLQPLKPSPPSNDLIYSAFTSDGALSVPVSLSPAGOGIKOP 1973

RESULT 10
US-09-854-856-62
; Sequence 62, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walker, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.

Db 5323 SMQTLRLVNNQNAIHSTNSYFNEDSTQKNTYDNDNGSTYITGQHNPELNKSTIDOTI 5382
Qy 2111 HQDKTLPPAQSSVGPAAKQDTAQPSPQPOQPOSPAQPEVOTQPEVOTQTV--SS 2168
Db 5383 SRINT--AKNDLHGVEKLQDKG-----TANQIEIGQLGYNLPQKSGEESLVNGS 5430
Qy 2169 HVPSEAPHTAQSCKPQVAQSPQSNVQSGSPRVQSPQSPQTRIRPSTPQSPQSQSQV 2228
Db 5431 NTRSEVE-EHLEAKSLNNAMKQURDKVAETKNYK---QSSDYINDSTEHGQYDQALQE 5486
Qy 2229 QTTTSOPIPIQHTSLQIPSGQPQ-SOPQVQSSQTQLSSQQT----- 2270
Db 5487 AENIINEI-----GNPLNKSEIEQKLQQLTDAQNALQSHLLEAKNNAIT 5533
Qy 2271 -LNOVSVSSPSPRPOLOQPOQPVIAVPQLOQOVVLQSQIOS--QVVAQIAQAGSGVPQ 2327
Db 5534 GINKLTALNDAQROKAIENVAQAOQ-TIPAVNQOULTDREINTAMQALRDVKVGOQNNVHQ 5592
Qy 2328 I-----KLOLPQIQQSSAVQTHQIQNVVTVQAAASVQEQ 2362
Db 5593 SNYNEDEQPKHNDVNSVQAGOTTIDKLDQDPMKNKEIQAININTQOT--ALUSGENKL 5650
Qy 2363 QRVOQLRDQO---QKKKQOOIEIKREHTLQASNOSEIIQKVVMKH-NAVIEHLKO--- 2414
Db 5651 HTDQESTNRQIEGLSSLTAQINAEKDLVNQAKTRTDVQAKLAAAKAINSAMSRLDGIQ 5710
Qy 2415 -----KKS-----MTPAEREENORMIVCQVMKYILDKTDKEEQAARKRRE-- 2457
Db 5711 NKEDIKSSAYINADPTKVTAIDQALQNAENIINATPNVELNKATIEQALSrvQOAAQODL 5770
Qy 2458 ESVEQ-KRSKONATK---LSALLFKHKEQLRAEILKRALLDKLDQIEVQEEELKRDKI 2512
Db 5771 DGVQOLANAKQOATQTVNGLSLNDGQKRELNL-----LINSANTRTKVQEEELNKATEL 5824
Qy 2513 KKEKDLM-----QLAQATAVAAPCPP-----VTPVLP----- 2539
Db 5825 NHAMEALRNSVQNDVQKSSNVYNEQPEQHNVDNAVNEAQATINNNAQPVLDKLAIER 5884
Qy 2540 -----APAPPSPPPPGVQHTGLLSTPTLPVSAQKREKREKSSSKS 2595
Db 5885 LTQTVNTTKDALHGAQKLTQDQAAETGIR--GLTS-----LNEPQKNAEVAKVTAATT 5936
Qy 2586 KKKMISTSKETKDTKLYCICKTPYDE-----SKFYIGDRCONWYHGRVCGILOSEAE 2640
Db 5937 RDEVNTRQEAATLDTAMLGLRSIKDKNDTKNSKSKYINEDHDOQOQAYDNVANNQA--Q 5993
Qy 2641 LIDYVCPQCQSTEDAMTVTLPLT-----EKDYEGLRVLRSLQAHKMAWPF 2687
Db 5994 VIDE---TQATLSSDTINQLANAVTOAKSNLHGDTKLQHKDSAKQTIQALQNLNSAQKH 6050
Qy 2688 LEPVDPNDAPDYGVIVKEPMDLATMEERVQRRYE-----KLTEFVADMTKIIFDNC 2738
Db 6051 ME-----DSLIDNESTRQVQHDLTEAQALDGLMGALAKESIKDNTNIVSNG 6096
Qy 2739 RYVN--PSDSPFYQCA 2752
Db 6097 NYINAEPSKKQAYDAA 6112

RESULT 8
US-09-854-856-48
; Sequence 48, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Waikie, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE REFERENCE: Encoding the Same
; CURRENT APPLICATION NUMBER: LEX-0178-USA
; CURRENT FILING DATE: 2001-05-14

; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1939)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-48

Query Match 2.5%; Score 358.5; DB 4; Length 1939;
Best Local Similarity 18.2%; Pred. No. 9.3e-13;
Matches 415; Conservative 292; Mismatches 123; Indels 847; Gaps 95;

Qy 399 TNKARGSNKFLAANAEIILESIRAKKGDIDNVKSPETEKDKNETENDSKDAENREEF 458
Db 100 TSKDRPVQSPSLVSGSKKEPPPPARGSGGG-----SAKEPQERSQQQDDI-----EEL 147
Qy 459 EDQSLEKDSDDKTDDDDPEQCKSEVGFQKSEKSNGLSESPCAGKAGSGSTRI-ITRLRN 517
Db 148 ETKAVGMSNDGRFLKFDIEIGR---GSFKT-----VYKGLDTTETVEVANCE 192
Qy 518 PDKSLQSLKSOQVAAAAHEANKLFKEKEVLVNVQSGEISRLSTKKEVIMKGN--INNYF 575
Db 193 QDRKLTKSEQR-----FKEBAEMLKGLQHPNIVRFYDSEWSTVKGKCCIVLT 241
Qy 576 KLQEGEKYRVYHNOYSTNSFALNKHQRED-----HDKRRHLAHK-----FCLTPA 621
Db 242 ELMTSGTLKTYLKRKFYVNMKTVLRSMWCRLKGLQFLHTRPTPIIHRDLKODNIFITGPT 301
Qy 622 GEFKWSGVHSGKVLTTISTLRLTITOLENNIPSSFLHPNWSHRANWIKAV----- 672
Db 302 GSVKIGD-----LGLATLK-----RASFAKSVIGTPEPMAP 332
Qy 673 QMCKPRE-----FALALALECA-----VKPV-----VMLPIW 701
Db 333 EMYEEKYDESVDVYAFGCMCMLEMATSEYPYSECNAAQIYRVVTSVGVKPFASFDKVAIPEV 392
Qy 702 RE-----FLGTHLRHWTISI-----EREKEK-----VKKK 727
Db 393 KEITEGCIROKNDERYSIKDLLNHAFFQEEFGVRAVEAEDDEGKIAIKMLRIEDIKKL 452
Qy 728 EKQOEETMQAATWVYTPPVKQVWVKQKEEYRVTVG---GCWSWTSKTHVYRFVPKL 784
Db 453 KGKYKDNEA-----IEFSFDLERDVPDVAQEMVESGYVCEGDHKTWAKA----- 497
Qy 785 PGNTNVNVRKSLGCTKNMNDENMDESCKRSPKKIKIEPDSEKDEVKGSDAAGADQ 844
Db 498 -----IKDRVSLIKRKRQRLVREEQEK-----KQESSLKQOQVESSASQ----- 540
Qy 845 NEMDISKITEKKDOOVKELDDSDSKPCKEPEPNEVDMDKMTESHVNCQESSQDVVVNSE 904
Db 541 -----TGIKQLPSASTGIP-----TASTTSASVSTQVEPEPEPA 574
Qy 905 GFHLRTSYKKTKSSKLDGLLERRIKOFTLEEKORLEKIKLEGIGIKGTSTNKKSLNS 964
Db 575 DQHQOQLYQOPSISVLSGDIVD-----SQGSSSVFTESRVSS 611
Qy 965 ESPVITKAKGCGQSD-----SMRQESPNNANDQPEDLIQCGSQSSSVLRM 1011
Db 612 QQTV-----SYGSGHEQAHSCTGTPGHIPSTVQAQSPGHVYPPPSVAQSQGQSPSSSL 667
Qy 1012 SDPSHTTNKLYPKDRVLDDVSIKSPETKCPKQNSIENDI-----EEKVSDLASRQOEP 1064
Db 668 TGVSSSOPIQHPQOQ-----QGIQQTAPPQOQTVYLSQTSSTSEATTAPQVSPQAP 720
Qy 1065 T-----KSKTKGNDFFIDDSKLASADDITGLICKNKKPLQIOEESDTIVSSSKSLAHS 1116
Db 721 QVLPQVSAGKQSTQG-----VSQVAPAEPAV-----AQP--QATQPTTLASSVDSAH 767

QY 1712 PRETP--TPQKGLRSSALRPKRPETPKOTGPV-----IETWVABE----- 1751
Db 1270 IASTPLPQAAGTSTPLLPQVPSIPPLVQPVANVAVQOTLIHSQPQALLNQPHTHC 1329
QY 1752 -ELE-----LWEIRAFARVE---KEKAQAVEQAKKLEQO---KPTV---IATS 1792
Db 1330 PEVDSOTOPKAPGIDDIKTLEKRLSLFSEHSSGAHASVSLETSLSVIESTVTPIGPTT 1389
QY 1793 TTSPT---SSTSTISPAKV---WVAPISGVTTGKWLTKVGVSPATVTFQKNKF 1845
Db 1390 AVAPSKLLTSTCLPPTNLPLGTVALPVTPVTPG-----QVSTPVST----- 1435
QY 1846 HOTFATWKGQSGSVQVQOQKVLGIIPSTSTGTSQOTFTSPQRTATVTRPNTSGSG 1905
Db 1436 -----TSGVPGTAPSKRPPLTKAPVLPVGTLPAG 1465
QY 1906 TTSNQVITGPQIRGMVTRTPLOO-----SILGKAILIPTVMVQPG 1948
Db 1466 TLPSEQLPPFP--GFSLTQSOQPLEDLAQLRRTLSPEXITVTSVAGPVMAAPTATEA 1523
QY 1949 A--POQVMTQIIRGQPVSTAVSA-----PNTVSTPGQKS-----LTSAT 1986
Db 1524 GTQPKGVSVQKGVPLATSSGAGYFKMGRFOVSVAAQGAQKGNKSEDAKSVHFEST 1583
QY 1987 STSNTOSSASQPPRQOQGVKLTMALQTLQTOGHGNGOGLTVVIOGGQTTGOLQIPOG 2046
Db 1584 SESSVLSSS-----PESTLVKPE-----PNG 1605
QY 2047 VTVLPGQOQLMQAAMPNGTVQRFLETPLATTTASTTT-----TTVSTTAAGTCE 2098
Db 1606 ITI--PGISSDVPESA-----KTTASEAKSDTGQPTKVGRFQVTTTANKVGR 1651
QY 2099 QROSK-----LSPQMVOHQDKTLPPAQSSSVGPAKAQAPQTAQPSARPQOT 2144
Db 1652 FSVSKTEDKITDKKEGVPVSPFMDLEQAVLP---AVIPKKEPSELSESHLNGPSS 1706
QY 2145 QPO-----SPAQP----- 2152
Db 1707 DPEAFLSRDVGSGSPHSQLSKSLPSONLSQSNFNSVMSYMSDNESDIEDL 1766
QY 2153 -----EQV-----TOPEVOTQTTVSSHVP-----SEAQPHAS 2181
Db 1767 KLELRRLDKHLKEIQDLSQRKHETESLYTKLGKVPVAVIIPPAAPLSGRRRPTKSG 1826
QY 2182 SKPQVAA---SQSQ---SNVQSQPVVRQSPSQTRIPSTPSQLSGQSQVQTTSQPI 2236
Db 1827 SKSSRSSSLGNKSPQLSGNLGSAASVLPQQT-LHP--PGNIPESGQNL---LQPL 1879
QY 2237 PIQPH-----SLQIPSQOQOPQOQVQSQSTQTLSSGQTLNQVSVSPSRPQLQ 2285
Db 1880 KPSPSSDNLYSFTSDGAI SVPSLSAP---GQGTSTNTV--GATVNS-----Q 1923
QY 2286 IOQPOP 2291
Db 1924 AAQAP 1929

RESULT 6
US-09-854-856-14
; Sequence 14, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19

; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 2136
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2136)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-09-854-856-14

Query Match 2.5%; Score 364.5; DB 4; Length 2136;
Best Local Similarity 18.1%; Pred. No. 4.7e-13;
Matches 414; Conservative 298; Mismatches 725; Indels 849; Gaps 96;

QY 399 TNKAGSNKSFLLAANEELIESIRAKKGDIDNVKPEETEKOKNETENDSKDAENREF 458
Db 160 TSKDRPVSPQSLVSGKEEPPARSGGG-----SAKEPQERSQOODDI-----EEL 207
QY 459 EQSLEKSDDDKTPDDDPQKSEVGDFFKSEKSNSELSGAGKAGSSTRI--ITRLRN 517
Db 208 ETKAVGMSNDGRFLKFDIEIGR---GSFKT-----VYKGLDTETTVAVACEL 252
QY 518 PDSKLSQLKSQVAAAHEANKLFEGKEVLVVVNSOGEISRLSTKKEVIMKN--INNIF 575
Db 253 QDRKLTKSERQ-----FKEEAEMLKGLQHPNIVRFYDSWESTVKGKICIVLVT 301
QY 576 KLGQEGKYRVYHNOYSTNSFALNKHQRED-----HDKRRHLAHK-----FCLTPA 621
Db 302 ELMTSTGLTYLKRKVMKIKVLRWCROILKGLQFLHTRTPPIHRDLKCDNIFITGPT 361
QY 622 GEFKNGSVHSGKVLITISTRLTITQLENNIPSSFLHPNASHANWIKAV----- 672
Db 362 GSVKIGD-----LGLATLK-----RASFAKSVLGTPEFMAP 392
QY 673 QMCSKPRE-----FALAIALECA-----VKPV---VMLPTW 701
Db 393 EMVEEKYDSDVLYAFGMCMLMATSEYPYSECQNAAYIRKVTSGVKPASFDKVAIEV 452
QY 702 RE-----FLGHTRLHRMTST-----EREKEK-----VKKK 727
Db 453 KELIEGCIQNDKDERYSIKDLLNHAFFQETGVRVELAEEDDECEKAIKLWLRIEDIKKL 512
QY 728 EKKQEEETMQATWVKYTFPVKHQVWKQGEERYVTGY---GGWSWISKTHYVRVPKL 784
Db 513 KGKYKDNEA-----IEFSFDLERDVPDVAOEMVESGYVCEGDHKTAKA----- 557
QY 785 PGNTVNYRKSLGCTKNNMDENWDESDKSKSPKIKIEPDSEKDEYKGSDAAGADQ 844
Db 558 -----IKDRVSLIKRREORQLVREPEKK-----KOEESLKKOQVEQSASQ----- 600
QY 845 NEMDISKITEKKDQDVKELLDSDDPKCEPEMEVDMDMKTESHVNCQESSQVDVYVNSE 904
Db 601 -----TGIKQLPSASTGIP-----TASTTSASVSTQVEPEPEA 634
QY 905 GFHLRTSYKKTKSKLDGLLERRIKOFTLEEKORLEKIKLEGIGIKGTSTNSKNLS 964
Db 635 DQHQQLQYOQOOPSISVLSDDGTV-----SGQSSSVFTESRVSS 671
QY 965 ESPVITKAKEGQSD-----SMRQEQSPNANDQPEDLIQCSQSDSVLRL 1011
Db 672 QQTV-----SYGSOHQAHSTGTVPGHIPSTVQASQPHGYVPSSVAQSQSQPSSSL 727
QY 1012 SDPSHTNKLKPKDRVLDDVSRKSPETKCPKQNSIENDI-----EEKVSDIASRGQEP 1064
Db 728 TGVSSSQPTQHPQQO-----QGIQOTAPQOTVQVLSQTSSEATTAQVSPQAP 780
QY 1065 T-----KSKTKGNDFFIDDSKLASADDIGTLCKNKKPLIQEESDITVSSSKSALHS 1116
Db 781 QVLPQVVSAGKQSTQG-----VSQVAPAEFVAV-----AQPTTTLASSVSDSAHS 827
QY 1117 SVPKSTNDRDATPLSRAMDFEK-----LGCDSSENSTL 1150

QY 2607 CKTPYDESKFYIGCDRCQNNYHRCVGILOSEAELIDEYVCPQCSTEDAMTVLTPLTEK 2666
Db 2306 LET-----QIEC-----LMSDQECVK-----RREEEIEQLNEVIEK 2337
QY 2667 DYGLKRV--LRSLQAKMAWPLEPVDNDAPDYGVIKEPMDLATMEERVQRRYYEKL 2724
Db 2338 LQELANIGOKTSMNAHLS-----BEADS-----LKHQLDVIAEKLALEQOQVETA 2384
QY 2725 TEVADMTKIFDNCRYNPNPSPFYQCAEVLESFFVOKLKGFKFASRHHNNKLOS 2778
Db 2385 NEEMTEPMKNVLKETNF-----KMQLTQELFSLKKRESVEKIQS 2424

RESULT 5
US-09-854-856-46
; Sequence 46, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 2076
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2076)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-46

Query Match 2.5%; Score 364.5; DB 4; Length 2076;

Best Local Similarity 18.1%; Pred. No. 4.5e-13;
Matches 414; Conservative 298; Mismatches 725; Indels 849; Gaps 96;

QY 399 TNKARGSNKSFLLAAANEIELESIRAKKGQIDNVKSPETEEDKDNETENDSKDAEKNREEF 458
Db 100 TSKDRPVSPSLVSGSKSEPPPARSGGG-----SAKEPQERSQQDDI-----EEL 147
QY 459 EDQSLKSDDKTPDDPEQKSEVGVDFKSEKSNGLSESPGAKGASGSTRI-ITRLRN 517
Db 148 ETRAVGMSNDGRFLKFDIEIGR---GSFET-----VYKGLDTEETVEVAVCEL 192
QY 518 PDSKLSQLKSQVAAAHAEANKLFKECKEVLVNVNQSQETISRLTKKEVIMKGN--INNYF 575
Db 193 QDRKLTKSEKOR-----FKEEAEMKLGQHPNIVRFYDSWESTVKGKCKIVLYT 241
QY 576 KLQGEQKIRYVHNOYSTNSFALANKHQHRED-----HDKRRHLAHK-----FCLTPA 621
Db 242 ELMTSGTLKTYLRFKRVKMKIKVLRWCROILKGLQFLHTPTPIIHRDLKCDNIFITGPT 301
QY 622 GEFKNGSVHGSKVLITSTLRLTITQLENNIPSSFLHPNWSHRANWIKAV-----672
Db 302 GSVKIGD-----LGLATLK-----RASFAKSVIGTPEFMAP 332
QY 673 QMSCKPRE-----FALALATLECA-----VKPV-----VMLPIW 701
Db 333 EMEEYKESVDVYAFGCMCMLEMATSEYPYSECQAAQIYRRVTSGYKPAFDKVAIPEV 392
QY 702 RE-----FLGHLRHRMTSI-----EREEXE-----VKKK 727
Db 393 KEIEGCIRQNKDERYSIKDLNHAFFQETGVRVELAEDDGEKTAIKLWLRIEDIKKL 452

QY 728 EKKQEBEETMOQATWVKYTPPVKHQVWKQGEERYVTGY---CGWSHISKTHYRVFVKPL 784
Db 453 KGKYKONEA-----IEFSFDLERDVPEDVAQMWESGYVCEGDHKTMAKA-----497
QY 785 PGNTNINRYKSLBGTNNMNDENDESKRCKSRPKKIKLEPDEKDEKGVGDAAGADQ 844
Db 498 -----IKDRVSLIKRREKQQLVREGEKK-----KQEESSLKQOVQSSASQ-----540
QY 845 NEMDISKITEKKDDQVKELDSDSKPCKEEPMEVDDDMKTESHVNCQESSQVDVNVVSE 904
Db 541 -----TGIKQLPFASTGIP-----TASTTSASVSTOVEPEPEA 574
QY 905 GFHLRTSYKKTAKSKLDGLLERRIKQFTLEEKORLEKIKLEGIGKIGTSTNSSKNLS 964
Db 575 DQHQQLQYQQPSISVLSLDTVD-----SGQSSSVFTESRVSS 611
QY 965 ESPVITKAKEGCQSD-----SMROBOSPNANNQDQEDLIQGCSDSSVLRM 1011
Db 612 QQTV-----SYGSHQEAHSTGTVPGHIPSTVQAQSQPHGVYPPSSVAQGSQSQPSSSL 667
QY 1012 SDPSHTTNKLYPKDRVLDDVSIKSPETKCPKQNSIENDI-----EEKYSDLASRQEP 1064
Db 668 TGVSSSQPIQHQQQ-----QGIQQTAPPOQTVOYLSQTSSTSEATTAAQVPVSOQAP 720
QY 1065 T-----KSKTKGNDFIDDSKLASADDIGLICNKNKAPLIOBESDFTVSSKSKSALHS 1116
Db 721 QVLPOVSAGKQSTQG-----VSQVAPAEPAV-----AQP--QATQPTTTLASSVDSAH 767
QY 1117 SVPKSTNDRDATPLSRAMDFEGK-----LGCDSSESNSTL 1150
Db 768 DVASGMSDGNENVPSSSGRHEGRTTKRHYKRSVRSRHEKTSRPKLRLNVNKGDRVV 827
QY 1151 ENSSDT-----VSIG-----DSSEE-----DMIVQNSNESISEQFRTREQDVEYLEP 1192
Db 828 ECQLETHNRKMVTFKFDLDGDNPEEJATIMVNDFFILAIERESFVDQVR---ETIEKADE 884
QY 1193 LKCELVSGETGNCEDRLPVKGTANGKKPSQOKKLEER---PVNKCSDQIKLKNITDKK 1249
Db 885 MLSEDSVSVEPEGD--OGLSLOQKDDYGFSGSGLGEFQKQPIPASSMPOQIGIPT-----938
QY 1250 NENRESEKKGQRTSTFOINGKDNKPKIYLKGE--CLKEISESRVSVGNVEPKVNNINKI 1307
Db 939 -----SSLTQV-----VHSAGRRFVSPVPSRRLRESKVFPP-----969
QY 1308 IPENDIKSLTVKESAIRPFINGDVIMEDFNERNSETKSHLLSSSDAEG-----NYRDS 1361
Db 970 ---SEITDTVAASTAQSPGN--LSHSASSLSLQQAFLSRAQMTGEGPNTAPPNFSHT 1023
QY 1362 LETLPSTK---ESDSTQTTTPSACPSNSVNOVEDMEIETSEVKKVTSPITSEESNL 1418
Db 1024 GPTFPVVPVPLSSIAGVPTTAAATAPVPATSPSPND--1STS---VIOSEVTVPTEGI 1077
QY 1419 ---SNDFTDENGFLPINKNENGVESKRKTIVITEVTTMTSTVATESKTVIKYKGGKQOT 1473
Db 1078 AGVATSTGVVTSGLPI---PPVSESPVLSVSVSSIIIPAVVSIISTSPSLQVPTSTSEI 1134
QY 1474 VVSTENCACKSTVTTTTTTTKLSTPSTGGVDIIISVKEQSKTVVITVTTDSLTSTGGTL 1533
Db 1135 VVSSTALYPSVTVSATSA-----SAGGSTATPGKPPA--VVSQQAAGS--TTVGATL 1183
QY 1534 VTSMTVSKEYSTRDKVKMLKFSRPKPTRSGTALPSYRKKFVTKSTKKSIFVLPNDLKLKA 1593
Db 1184 TSVSTTTTSPSTASQSLI-----QLSSSTSTPTLAETVVVSAH-----1221
QY 1594 RKGIREVPYFNKAPALDIMPVPSRPTFGITWRYRLQTVKSL--AGVSLMLRLWLAS 1651
Db 1222 -----SLDKTSHS---TTGLAFSLASFSSSSSPGAGVSSVIS-----1256
QY 1652 LRWDDMAAKVPPGGGSTRTETSETITTEIIRRDVGPYIRFEYICIRKIICPIGPVET 1711
Db 1257 -----QPGG-----LHPLVIPS 1269

[illegible]

RESULT 3
 US-09-134-001C-5080
 ; Sequence 5080, Application US/09134001C
 ; Patent NO. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 5080
 ; LENGTH: 3696
 ; TYPE: PRT
 ; ORGANISM: staphylococcus epidermidis
 US-09-134-001C-5080

	Query Match	2.98;	Score 410;	DB 4;	Length 3696;
	Best Local Similarity	17.7%;	Pred. No. 1.9e-15;		
	Matches 464;	Conservative 427;	Mismatches 1027;	Indels 704;	Gaps 108;
Qy	311	QKNKPYIRHPIGYD-PSRRKYWFLNRR--LIITEEDTENENEKKIWIYYSKTVQLAELIDC	367		
Db	569	QVNKDII---PSNTYLLASYNKYKKLKERATQVLDEETNNTP-----	606		
Qy	368	LDKDYWEAEACKILLENREETHRMIDTDLTNKARG-----SNKSFLAAANEELIESI	421		
Db	607	FNQRYSTQIDDLLLHELQTLINRVYSASREINDKAOQMTDAVYDSTELTTTEKDTLVDQI	666		
Qy	422	RAKGGDIDNVKSPE-----ETEKDKNETENSDSKAEKNREFFEDOSLE	464		
Db	667	ENHKNEILSNIDDELTDGQVERVKEAGHLTLESDTPHPVTKPNARQVNNRAD--QQRITL	725		

Qy	465	KSDDKTPDDDE-----QGKSEVGDGFKSEKSNGLSESPPCAGKSGASTRIITR	514
Dd	726	RNHEATTEBQEAIRQVEAHSADAKIGAEAFDTTVNEARD-----NGTKLIAYD	777
Qy	515	LRPDSKLSQLKSQQAAAAHAENKLFKEGKVLVVNSOGESIRLS-----TKKEVI	566
Dd	778	VPNPTK-----KAERAAVYVNSANSIKD-----INNNTQATLDERNDATALVNRSKDEAI	828
Qy	567	MKGNNIFYKLGOBKGYRVVHNOYSTNSF-----ALNKHQHREDHKRRHLAH	614
Dd	829	Q--NINT--AQGNDDYTEAQN--GNTIQOVPPLTPVKRONALATINAKADEQRKLIQAN	882
Qy	615	KFCULTPA---GEFKWGSV-HGSKVITISTLRITITOLENNIPSSFLHPNWASHRAWJK	670
Dd	883	NNATTECKADAERKVNEAVITANQONTNATTNRDVDAQOTT-----GSGIIS	929
Qy	671	AVQMCSPREFALAILCAVAPVVMPLPIWREFLGHTRLHRMTSTIEREKEKVKKKKK	730
Dd	930	AISPATKIKEDARA-AVEAKIAQNOOI-----NSNNMATTEEKEDALNQVEAIK	978
Qy	731	QEHEETMQQATVWKYTFPVKHQVWKOKGEEYRTVGYGWSWI SKTHVYFPVKLPQNTNV	790
Dd	979	QAATATINQA-----QSTQVQSE-----AKNNGINTI	1005
Qy	791	NYRKSLGTTNMNDENNDESCKRCSPKKIKIEPDSKDEKVGSDAAKAGDONEMDIS	850
Dd	1006	NQOPNAVKNKTNTILEQKGNPKS---AIAQTDPDATTEE-----KOBAVS	1049
Qy	851	KITEKDODVKEILLDSDKPCREEPEMEVDMDMKTESHVNCBSSQVVDVVNSEGFHLRT	910
Dd	1050	AVSQAVTNGTHINOANSNDVDQOE-LSNAEOILITHNVNVQKKPO-----ARQALI AKT	1103
Qy	911	SYKKTKSSKLDGLLERRIKOF-TLEEKORL-EKI-----KLEGGITKGIGKYS	956
Dd	1104	NEKOSAINSDNEGTRIEBKRAIOGLSNDAKNLADEQITOAAASNQVNDNALNIGINSKIQ	1163
Qy	957	TNSKNI LSESPVITKAKEGCOSDSMRQEQSPNANDOPEDLIOGCCQSODSVLRMSDP SH	1016
Dd	1164	TNFTK---KQOARDQVNOKFOEKAELUNSTPHATQEBQDALTRLQAKETALNDINQAO	1220
Qy	1017	TTNKLYPKDRVL-----DVDYSIRSPETKCPKONSIENTIEEVSDIASRGQBPTS	1067
Dd	1321	TNQNV---DTALTSGIONIQNTOVNPK---KQEAKTINDIVQOHKQSIONDDATTEE	1274
Qy	1068	KTGNDFPIDSKLASADDITGLICKNNKPLOESTDIYVSSKSAIHSSVPKSTNDRDA	1127
Dd	1275	KEVANN-----LVNASQ-QNVISKIDNATTNNQIDGIVSDGRQSI NATTPOTSIKRNA	1326
Qy	1128	TPLSRAMDFEKLGCDSFSNSTLENSDTSVIO---DSSMEDIMVONSNESIS EQPFTR E	1184
Dd	1327	-----KNIDIIKAADKKTIQORINDATDEE---IQANKKIEE-----	1361
Qy	1185	QDVEVLEPLKCYLGVESTGCEDLRPLVKGETANGK----KPSOOKKLEERP V--NKCS D	1238
Dd	1362	-----AKTEAKDNTOIRSTDOVNEAKTINGINKIENTIPATTVKSEARQAVONKANE	1413
Qy	1239	QIK-LKNITTDKKNHNRESKKQORTST-----FQINGKDNPKPI-YLKGECLKEIS--	1288
Dd	1414	QINHIIQTPDATNEKEQEI---NRVSAELARVQQA INAEHTTGQVKTIKDDAITSLIRS	1470
Qy	1289	-----ESRVGVNPEPVNNINKITI PNDIKSLTVKESAIRPFPI NGDIVIMEDFNERNSS E	1343
Dd	1471	NAQVVEKESARNALEQKATOQTQFINNDNNAIDEEKVA-----NNULVIA---TKQKSLD	1522
Qy	1344	TKSHLLSSDAE----GNYRDSLFTLPKTESDSTOTTPFSACPESNSYNQVEDMBE IET	1399
Dd	1523	NINSLSNNOVENAKVAGINEIANVLPAATAVAKSKAKKIDOKL---AQOINQIOHTQ TAT	1579
Qy	1400	SEVKVTSSPPTSEEENLSNDFIDBNGLPIKNNENVNGESKRKTIVITETMTSTVATE	1459
Dd	1580	TEEK-----BAAIQLANQKSNEARTAIONEHSNNGVAQAK-----	1614

US-09-134-001C-4463
; Sequence 4463, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 3.0%; Score 432.5; DB 4; Length 2137;
Best Local Similarity 17.2%; Pred. No. 3.8e-17;
Matches 395; Conservative 355; Mismatches 999; Indels 547; Gaps 77;

QY 394 ITEDLTNKGASNSFLAAANEET-----LESIRAKKGDIDNVKSPETEKD- 440
DB 149 MSSEVTNDQSEKAGISQOSETSNOSSELNTYASTDHSVETTTNNDNTAQODNKNSSN 208
QY 441 --KNETENDSKAEKNEEFEDQLEKSDDKTPDDPEQCKSVGDFKSKSGELSES 498
DB 209 VTSKSTQSTSSSEKNTSSNLTQSIETKATDSLATSAPRTNOISNLTSTSTNQSSPT 268
QY 499 PGACKGASGPRITIRLNP-----DSKLSQLKSOVAAAHEANKLF-----KEG 544
DB 269 SFANLRTFSRPTVLTMAAPTSTTTTSSLTSSVNVKDNFNEHNNLSSGATYDPKGT 328
QY 545 KEVL--VVSQGEIS---RLSTKEVIMKGNIN--NYF-----575
DB 329 IATLTPDAYSQGAISLNRDLSDNRSFRTGKVLGNRYEGYSPDGVGGDGFAGFSPG 388
QY 576 ---KLQEG-----KYRVVHNOYSTNSPALNKHQHRE-----D 605
DB 389 PLGOIGREGAAGVIGGLNNAFGKLDYHNTSTPKSAKADARNVGGGAGFAFVSTD 448
QY 606 HD-----KRRHLAHKFLTPAGE-----FKWNGSVHGSKVLTISTRLTITQLENNIPSS 655
DB 449 RNMATTEASSAALKNVQPTDNSQDFVIDYNGD---TKVMTVTYAGQTFR-----497
QY 656 FLHPNASHRANWIKAVQMCSPREFALALAILCAVKKPVVMLPIWREFLGHTRLRHMTS 715
DB 498 -----NLTDWIKN---SGGTTFLSMTASTGGAKNLQOVQFGTFEYTESAVAKVRY 545
QY 716 IEREE-KEKVKKKEKQEEETMQOATWVKYTFPVKHQV--WKOKGEEYRVT-----GY 766
DB 546 VDANTGRDIIPPKTIAGEVDATV-----IDKQLNLLKNSGYSYVSTDALQNSNY 595
QY 767 GGWSNI-----SKTHVRFVFKPLPCNTNWNVYKSLGKTKNNMDNMDSDKRCRSR 818
DB 596 SETSGTPTLKLNTSSQTVYKF-----KDVQGPQISVD-----SQTRVEGKT 637
QY 819 PKKIKI-EPDSEKD-----EVKGSAAKGAQD-----NEMDISKITEKKDDQVRELL 864
DB 638 INPITITTDNSKDVLTATTVTGLPSGLSFQDTTNTITGTPSEVGTITVT-----VNTT 690
QY 865 DSDSDKCKEPMEDVDDMDKTESHVNCQESSOV-----DVNVNS 903
DB 691 DATGNVTSKQFTIITQITISPVNVNTPSQASEVFPINPITITATDNGSKVTVITVGLP 750
QY 904 EGFHLRTSYKKTKSSKLDGLLERRIKOFTLEEKORLEKIKLEGKIGKIGKTSTNSSKNL 963
DB 751 QGLKFDASTNSIVGTPQIGTNTITIESTDASGNKTTKINYE-----VTRNSASDS 802

QY 964 SESPVITKAKEGCO-----SDSMRQEQSPNANNOPEDLIOGCSQSDSVLRMSDPSTHT 1018
DB 803 TSTIVNSVSTSIENSTSLSDSVKASQSLSTSKLSLSE--SASTNSTSIQASASASTS 860
QY 1019 NKLYPKDRVLDVDSIRSPTKCPKONSTENDIEEKVSDLASRQBPETKSKTKGNDOFFIDD 1078
DB 861 KQL-----SEASTSTSDASAKSESTSKASTS-----LSE 893
QY 1079 SKLASADDIGTLICKNKKPLIQEESDTIVSSSKSALHSSVPKSTNDRDATPLSRAMDFEG 1138
DB 894 STSTSVSDSASVSTSES---ASTSTSVSGSTSTSIDSTSTSDSASIKASESAS--TS 948
QY 1139 KLGCDSSNSTLENSSDTIVSIQDS--SEEDMIVQNSNISIOFRTREDQVLEPLKCEL 1197
DB 949 KLLSESVSTSTSDASTSTSVSDNSASTSKSTSTSVSDSTSTSD-----997
QY 1198 VSGESTGNCEDRLPVKGTGANGKKPSQOKKLEERPVKNCSDQIKLKNNTDKKNENRESE 1257
DB 998 --SASTSTSE-----SESDSASTSLESTSTSVSDSTSTSTSDSASMSASE 1043
QY 1258 KKGQRTSTFQINGKDNPKIYLKGECLKEISESRVVGNGVPPKVNINKIIPENDIKSLT 1317
DB 1044 SNSKSTS-----LSE-STSTSLSGSTSAST-----SDSASTS 1074
QY 1318 VKESAIRPFINGDIVMEDFERNSESTKSHLLSSDAEGNYRDSLETILPSTKESDSTQTT 1377
DB 1075 TSESE-----SDSTSTSLSESTSTSLSGS--TSASTSDSAST--STSESDSTSES 1120
QY 1378 TPSASCPESNVNOVEDMEIETSEVKVKTSPITSEESN-----LSNDFIDENGL 1428
DB 1121 T---SLESSTSVSDSTSASTSE---SASTSTSESNASASTSLSGSLSTSDSTST 1173
QY 1429 PINKNNVN--GESKRKVTITEVTTMTSTVAATESKTVIKVEKGDQTVVVSFENCACSPVT 1487
DB 1174 STDSASTSESESDSTSTSLSESTSTSLSDSTSTSEASASTSESDSTSESTSLSE 1233
QY 1488 TTTTIVTKLSTPSTGGSVDI---ISVKROSTV---TTVTDSLTITGGTTLVTSMVS 1540
DB 1234 STSTSVSDSTSDSASTSTSVSDSEASTSISELSSTSVSDS--TSTSTSDSASTST 1292
QY 1541 KEYSTRDKVKLMKFSRPPKTRSGTALPSRYKRFVTKSKTSIFVLPNDLKLARKGGIRE 1600
DB 1293 ESDSTSESTSLSEISTSVSDSTSA--STDSASTSTSES-----ESDSASTSLSG----1341
QY 1601 VPYFNNAKPALDIWPPSPRPTFGITWRYRLQTVKSLAGVSLMLRLWLASLRDDMAAK 1660
DB 1342 -----STSTSLS-----1348
QY 1661 VPPGGGSTRTETSETELTTTEIIKKRRDVGPGYRFEYCIRKIIICPIGVPTPKPTPQR 1720
DB 1349 -----DSTSTSTSDSASTSTSE-----ESDSERASTSL 1375
QY 1721 KGLRSSALRPKRPTPKQTGPV-----IETWVAEELELWEIRAFAR 1764
DB 1376 SGSTSTSLSDSTSTSDSASTSTSVSDNSASTSLSGSLSTSVSDSTSTSDSAST 1435
QY 1765 VEKEKAAVEOQAKRLEQOKPTVIATSTTSTSTSTISPAOKVMVAPISGVVTGCTK 1824
DB 1436 SED-----SBRASTSLSGSTSTSIDSTSTSTSDASTSTSVSE-----SNSTSTSI 1484
QY 1825 MVLTATKVGSPATVTFQOKNHFQTFATWVKQGSNGSVVQVQKVLGIIPSTGTSOOTF 1884
DB 1485 ESLSTSV--SDSTST-----STSDSASTSTSVSDSDASTSTSSSESV-----STSDSEST 1532
QY 1885 TSFQPRATVPI-----RPNTSGGGTTSNQVITGPQIRPGMTVIRTPLOQSTLGRKAI 1939
DB 1533 TSTSDSASTSTSVSESNSTSTSLSGSTSTSVSDSTSTSDSAS---ASTSESDSDAST 1589
QY 1940 RTPVMQPGAPQVMTQIIRQOPVSTAVSAPNTVSSTPGQK---SLTSATSTSIQSSAS 1996
DB 1590 SSSESVSTSVSDSTSASTSEASASTSTSVSDNSASTSLSESTSTSLSDSTSMSTSDAST 1649
QY 1997 QPPRPQOGVKLTMAQLTQLTQGGHNGGLVTVLQGGQQTTCGQLQLIPQGVTVLPGPQQ 2056

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	982.5	6.9	238	4	US-09-257-179-80	Sequence 80, Appl
2	432.5	3.0	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
3	410	2.9	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
4	373	2.6	3878	4	US-09-914-259-11	Sequence 11, Appl
5	364.5	2.5	2076	4	US-09-854-856-46	Sequence 46, Appl
6	364.5	2.5	2136	4	US-09-854-856-14	Sequence 14, Appl
7	363.5	2.5	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
8	338.5	2.5	1939	4	US-09-854-856-48	Sequence 48, Appl
9	338.5	2.5	1999	4	US-09-854-856-16	Sequence 16, Appl
10	358.5	2.5	2048	4	US-09-854-856-62	Sequence 62, Appl
11	358.5	2.5	2108	4	US-09-854-856-30	Sequence 30, Appl
12	353	2.5	2035	1	US-08-046-585-5	Sequence 5, Appl
13	353	2.5	2035	1	US-08-393-703-5	Sequence 5, Appl
14	353	2.5	2035	5	PCT-US93-11721-5	Sequence 5, Appl
15	352.5	2.5	1911	4	US-09-854-856-64	Sequence 64, Appl
16	352.5	2.5	1971	4	US-09-854-856-32	Sequence 32, Appl
17	352	2.5	2141	4	US-09-854-856-56	Sequence 56, Appl
18	352	2.5	2201	4	US-09-854-856-24	Sequence 24, Appl
19	347	2.4	2169	4	US-09-854-856-40	Sequence 40, Appl
20	347	2.4	2229	4	US-09-854-856-8	Sequence 8, Appl
21	346	2.4	2004	4	US-09-854-856-58	Sequence 58, Appl
22	346	2.4	2064	4	US-09-854-856-26	Sequence 26, Appl
23	346	2.4	8991	4	US-08-714-741-32	Sequence 32, Appl
24	341	2.4	2032	4	US-09-854-856-42	Sequence 42, Appl
25	341	2.4	2092	4	US-09-854-856-10	Sequence 10, Appl
26	338.5	2.4	2973	2	US-08-821-355A-7	Sequence 7, Appl
27	338.5	2.4	2973	2	US-09-003-687A-7	Sequence 7, Appl

THIS PAGE BLANK (USPTO)


```
Db 12354 QEVSVKETPTDISDRKPKTEKATSIILSEQESISVQEVSVKEAPGCTVETAKPKTEQATSI 12413
Qy 109 AIYEVLRNFTGTVLRSPFREDFCAALVSOEQCT-LMAEMHVLLKAVLREEDTSNTTFG 167
Db 12414 SPHEISIQEISVKEAP---TDICDKKPKSEKATSIILSEQESIAVEQSVKE-----A 12463
Qy 168 PADLKOSVNTLFLFDGWTWPEVLRYVYCESDKYHHVLPYQAEADPPYGPVENKIKVLQF 227
Db 12464 PGSTEEAKPT-----EYAKSSISPHESLTVOEV----- 12492
Qy 228 LVDFQTLTNIAREBELMSEGVIOYDDHCRVKHKLGLDILCCETCSAVYHLECVKPP--LEEV 285
Db 12493 -----LYKEASAEISDKKPKQFGKATSIILSEQESIAVEQSVKEAPGRVEDV 12538
Qy 286 -PEDEWQECVCAHVKPGVTVDCV-----AEIQNKPYIRHEPIGYDRSRKRYWFLNRR-- 337
Db 12539 KPKTEQATSIILSEQESITLSEQHSLTVQEVTVKGSABEISDKKP-----KTEKAISILSEQS 12588
Qy 338 LIITEEDTENENKWIYSTKVQLAELIDCLDKDYWEAELCKLILEMREIEHRHMDITED 397
Db 12589 ISVOEITVKEAPGSKVEDKLPQEA-----TSVISPLESLTVQEVSVKESGGD 12636
Qy 398 LTNKARGSNKSLAAANEIILE---SIRAKKGIDINVKSPEE-----TEK 439
Db 12637 ISDRRPKTERATSIIFSQESISVQEVSVKEAPGTVADVDPKPEEQATSIILSEQHSLTVQEV 12696
Qy 440 DKNETENDSKDAKNREE-----FEOQSL-----KSDDKTPDDDP---EQGKS----- 481
Db 12697 TVREIPTDISDKKPKSEKATSIILSEQESISVHEVSVKDPAGSMKDAKPKTEQATSVLSPH 12756
Qy 482 -----EVGDF--KSEKSNGLS-----ESPAGKAGASGSTR 511
Db 12757 ESITVQEVSVREPTIELSDKKPKSEKATSIILSEQESISVQEVSVKEAPGSMKDAKQKTEQ 12816
Qy 512 ITRLRNDSRLSOLKSOQVAAAHEANKLFKEGKVELVNVANSQGEISRLSKKEVIMK--- 568
Db 12817 ATSVISPHESITVQEVSVKEAPTELSDRPKKEEKATSIILSEQESIS---VQEVSVKEAP 12872
Qy 569 GNTNNPKLQEGEKYRYV---HNOYSTNSFALNKHQHREDHDKRRHLAHKFCLLPAGFEKW 626
Db 12873 GSVKD-LKLKTEQATSVISPHESITVQEVSV-KEAPTEISDKK-----PKTE--- 12917
Qy 627 NGSVHGSKVL-----TISTLRLTIT---QLENNIPSSFLHPNWS--HRANWIKAVQMC 677
Db 12918 -----QATSIILSEQESISQEVSVKAGPSLKDSEKQATSIILSEQHSLTVQEVTVKEA 12973
Qy 678 PREFALALALECAVKPVVMLPIWREPLGHLRLHRTSIEBKE-----KVKK----- 726
Db 12974 PADIS-----DVKE-----KTEKATSIILSEQESISVQEVSVKEAPGSM 13011
Qy 727 KKKQEBEE-----TMOQATWVKYTFPVKHQVWKQGEYRVTVGYGWSWISKT 775
Db 13012 KDAKPKTEQATSVISPHESITVQEVSV-VK---EVPTEISDKKPKSEKATSIILSEQESISV 13067
Qy 776 HVYRFYKPLPGNTNVNRYKSLGTX-----NNMDENMDESDKKCSRPKKIKTEPD 827
Db 13068 HEVS-VKESPGSMKDAKPKTEQATSVISPHESITVQEVSVKESPTIELSDKKPKKEKATSI 13126
Qy 828 -SEKDEYKGS-DAAKGADQNMEDISKITEKKDQDVKELL-DSDSKPKKEEPEVEVDMD- 883
Db 13127 LSEQESISIQEVSVKEVPESMKDLKPKTEQATSVISELLSILSVQEVSVKESPGDLSDKP 13186
Qy 884 KTE---SHVNCQESSQDVVNVSEGPHLRTSYKKTK--SSKLDGLLERIKOFTLEEQ 938
Db 13187 KTEKATSIILSEQESISVQEVSVKEAPGSEVEVPKTEQATSIISPHMSLTVQEVTVKE-- 13244
Qy 939 RLEKILEGKIGKGTSTNSKNLSRPVITRAKECCQSDSMROQSPNANDQPEDLI 998
Db 13245 -----VPADISDVKPKSEKATSIILSE-----QESISVQEVSVKEAPGSMKD----- 13285
Qy 999 QGCSQSDSVLRMSDPSTHTNKLKPKDRVLDDVYSIRSPETKCPKQNSIENDIEEKYSDLA 1058
Db 13286 -----AKPKTEQATSVISPHESITVQEVSVREVTEISDKK 13321
```

```
Db 1513 TEKTVE---ADDKFTTAVPLAGDEEESNLPKLPQDIFEEEA-PVAVTTAAPSKDD---GE 1565
QY 1440 SK-----RKTVITEYTMSTVATSKTVIKVEKGDQKQVVSSTENCAKSTVT 1487
Db 1566 QRPVEVEEPIEDGQKPIEDEESTPTS-----SENIEPESDRATTIAPSKEE--PSEPS 1618
QY 1488 TTTTWTKLSTSTGSGVDIISVKESQKVVVTTTVDLSLTGGLTIVTSMVTSKEYSTRD 1547
Db 1619 TGAPTKEPAEPSTDAPESEDES-KETPESEVPTTVAPA---GEKIPTSIIPTDEEPTAT 1673
QY 1548 KVKLMKFSRPPKTRSGTALPSYRKFKVTKSKTSIFVLPNDLKLARKGIGREVYFNVN 1607
Db 1674 SAPVAKPDVEDVEKETSTEIPTDAPASSEDEENS---STDQIPS-----EVP---E 1717
QY 1608 AKPALDIWYPSPRPFTGITWRYLQTVKSLAGVSLMLRLWASLRWDMAAKVPPGG-- 1665
Db 1718 KKP-----EVLAEIIPQSPETG 1732
QY 1666 -GSTRTETSETITTTTEILIKRRDVGPIRGIRFEYCIRKILICPIGVPTPKETPTPQRK-GL 1723
Db 1733 VGATAAPTTSDEVPPVQRLPE-----EVLAEIIPQSPETG 1768
QY 1724 RSSALRPKRPETPKQGPVVIETWVAEELELWEIRAFARVEK--EKAQAVEQ-----Q 1776
Db 1769 KQDETTAAAPSIDRK-----EPYVTEIDEEATTVAPISEKDEKPTTEBEKPVQKPTGEE 1822
QY 1777 AKRLQEQKPTVIATSTTSSTSTTISPAQKVMVAPISGSGVTTGKMKVLATKVGSPAT 1836
Db 1823 PSEEBEKRPIDQVSGTSPVSEAGSTES-----SEEVKPSTEGEVAEKPD-- 1873
QY 1837 VTFQKNKFNHQTATWVKQCGSNS-GVVQVQKVLGIIPSSTG---TSQQTFTSFOPRTA 1892
Db 1874 -----KQPSSTAQAPVETIPEISTELPAQDGDKPTSEAPVDSDEDTSA 1916
QY 1893 TVTIR-PNTSGSGCTTNSQVITGPOIRPGMIVIRTPLOQSLGKALITPVMVQGPAPQ 1951
Db 1917 PSDEKIPSVSG-----EEVEGPEV-----TTASPAQAAEDELKTPAESEPSSTD 1960
QY 1952 QV-MTOILIRGQPVSTAVSAPNTVSTSPGQKSLTSAUST-----SNIQSSASOPPRPQOG 2005
Db 1961 KVPETEYQKPEDTKADETPESVTVQSDVATSTSAVPAGDIEKDEQATASP--EEEEE 2018
QY 2006 VKLTMALQTLQTLGHGNGNGLVTVIQGQGTGQLQILIPQGVTVLP--GPGQOLMO-AAM 2062
Db 2019 IKPTIAPAAEIPQ-----PSEKEPVDE-QEVESGTKATPAESDGGQPIDEIAPA 2065
QY 2063 PNCVTQRFLETPLATTTATSTTTTSTVSTAAGTGEQKSLSPQMVHODK-----TL 2116
Db 2066 TSGPIDE-----ASTAAPTKEESTTVASAA-----SP--AVHDDIEIKDVTTTQ 2106
QY 2117 PPAQSSSVGPAKAQPOTA-----QPSAR-POPQTOPSPAQPEVQTOPQEVOT---QTTV 2166
Db 2107 PYADEKEVAAPQDETSTSDVSTDSPTAQDDEKQDKTEAPVAPTTVSSSTADSAADSSIP 2166
QY 2167 SSHVPS-----EAQPTH---AQSSKPOVA-AQSQPQSNVQSQSPVRV--QSPSQTRTRPS- 2215
Db 2167 TVEVPSPVEIDTKPMDDIMSQTITAPHTADGAASTSTDEDEQAPVTVSPQDAEKTPVSPAP 2226
QY 2216 -----TPSOLSPQSQSQVQVTTTQ-----PIPIOPHTSLQIPSGQPOPOPOVQSSQTQ 2264
Db 2227 QDSDKTPSSEAPQADAEIPATATPLDNDKNIPATVAPQTDGVPATAPALDEDKIQITTAAP 2286
QY 2265 L-----SSQTLNQVSVSPSPRQLOIQOQPOQVIAVPLQLOQVQVLSQIQSQVVAQIQ 2318
Db 2287 LDEEKIPSTAAPLDDDEKIPAPVSPVFDVPEPSSEKPAVSEYDGE-----ESTEPVHDVE 2341
QY 2319 AQSGVPOQIKLQLPQIQOQSSAVQTHQIONVVTVQAA-----SVOEQLRVQOQLRDQOQ 2373
Db 2342 TSTDEPTSDAKLKPPTSPATPSPSEPAEAEIYPETAAPLEKEVEPEKATEQPELEKETP 2401
QY 2374 KKQOQIETIKREHTLOASNOSET-----IQQVMKMHNAVTEHLKQKKSMT 2420
Db 2402 EKATEQPELEKETPEKATEQPELEKETPEKATEQPELEKETPEKATEQPESEVDEKTIPEP 2461
```

```
QY 2421 AER-----EENORMIVCNQVMKYILDKIDKBEKQAAKKRKRRESVEQRRSKQNA TKLSA 2474
Db 2462 VVKPSLDSTEEDE-----ESVESEESADKKDNKETEEDTDKKHEBEVPA 2508
QY 2475 L-----LFKHKEQLRAEILKRALLDKDLQIEVQEBELKRDLTIKKKEK 2516
Db 2509 VVSEIQPQSEEA VPTTGHPLFPH---LASSTTTPPAVDDR-----VGEE-----DEEN 2553
QY 2517 DLMQLAQATAVAAPCPVPT---PVLPAAPPSPPPPPGVQHTGLLSTPTLPVVASOKR 2572
Db 2554 TTVKLSSSTTTTSTPESVTSAPSTTTVASQQOQPIPTPPYG--HA-----PEYEDY 2603
QY 2573 KREEEKDSSSKKKKMISTTKETKDKTKLYCICKTPTYDESKFYIGCDRCQNMWYHGRV 2632
Db 2604 DEEEVPGTCRYAGKLYVSAQOIPRDDPCDFCF-----FRSDII 2644
QY 2633 GILQS-----EAEIDYVCPQCQ-----STEDAMTVITP 2662
Db 2645 CLQQSCPPPIAGCHERPISGFCPRVECPVMAAVLNITTTTSTTTTLP 2695

RESULT 15
ID Q95YM2 PRELIMINARY; PRT; 17352 AA.
AC Q95YM2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE I-connectin.
GN I-CON.
OS Procamburus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacidea; Cambaridae; Procamburus.
OX NCBI_TaxID=6728;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21423462; PubMed-11532946;
RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
RA Sun P., Maruyama K., Kimura S.;
RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RL sarcomeres of crayfish claw muscle.";
RL EMBO J. 20:4826-4835(2001).
CC -I- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB055861; BAB64297.1; -.
DR InterPro; IPR000577; FGGY_kin.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; ig; 48.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGC2; 13.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00933; FGGY_KINASES_1; 3.
DR PROSITE; PS00835; IG_LIKE; 50.
DR PROSITE; PS50002; SH3; 1.
KW Immunoglobulin domain; SH3 domain.
SQ SEQUENCE 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;

Query Match 4.0%; Score 572; DB 5; Length 17352;
Best Local Similarity 17.7%; Pred. No. 4.1e-15;
Matches 576; Conservative 570; Mismatches 1232; Indels 874; Gaps 133;

QY 3 SEEEEDGDAETQDSEDEDEMEDEDDDDSDSYPEMEDDDDDASYCT-----ESSFR 56
Db 12294 STEISDKPKSEKATSI LSEQESISQVSKDAPGSIKDAKPKTEQATSIISPESLTI 12353
QY 57 SHSTYSSTP---GRRKPRVHRPRSPILEE-----KQIPPLEFFPKSSDLMPVNEHIMNVI 108
```


RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminski J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Beran B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003808; AAM70936.1; .
DR FlyBase; FBgn0013988; Strn-Mick.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; Ig; 29.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00408; IGC2; 29.
DR SMART; SM00406; IGV; 4.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS50835; IG_LIKE; 26.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Immunoglobulin domain; Transferase.
SQ SEQUENCE 9270 AA; 1036749 MW; 0DD82D040B27C2AD CRC64;

Query Match 4.1%; Score 580.5; DB 5; Length 9270;
Best Local Similarity 18.2%; Pred. No. 8.2e-16;
Matches 574; Conservative 533; Mismatches 1160; Indels 885; Gaps 133;

QY 4 EEEE-----EEDGDAETQSDSEDEDEM-----EEDDDSDYPEEMDDDDASVCT- 51
DB 2614 EEELKASAKKQGDQIEQKSQKPEVSEVVAEKISEGKTEEPKPEEM-DTEAKSEKATV 2672
QY 52 -----ESSFRSHSYSTPGRRKPRV-----HPRSP----- 78
DB 2673 LDKQVLEKELEASAEKQGDQVEKKSQKPEVSEVVAEKISEETIEPKKPEVKDTEIKS 2732
QY 79 -----ILEEKDIPPLEFPKSSDLMVPNE--HIMNVIAIYELVRNGTVLRSLSPFR 127
DB 2733 EKATALDKQVLEKELEASAKQGDQVEKKSQKPEVSEIVA-----EKISEKT 2781
QY 128 FEDFCAALYSQEOCTLMAEMHHVLLKAVLREE--DTSNTTFGPADLKDSVNSFLYIDGM 185
DB 2782 IEPKPEVKDTE--TKSEKATALDKQVLEKELEASAKQGDQVEKKSQK----- 2831
QY 186 TWPEVLRVYCESDEKHHVLPYQEAEDYPYGPVENKIKVLQFLVDFLTNTIAREELMSE 245
DB 2832 --FEVSEVVAEKISE-----ETIEPKKPEVKETE-----KSEKATVLDK 2870
QY 246 GVTOYDDHRCVCHKLGLDCCETCSAYVHLECVKPPLEVEPDEWQCEV--CVAHVPGV 303
DB 2871 QVLEKELEASAKQGD-----QDVEKKFKQAEVSEVVAEKISEE 2910
QY 304 TDCVAEIQNKPIRHEPIGYDRSRKRYWFLNRLIIEEDTE-----NENEKKIWIYSTKV 359
DB | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

Db 2911 T--IEE--PKKPEVKDTEIKSEKATA-----LDKQVLEKELEASAKQGDQVEKKSQK 2962
QY 360 QLAELIDCLDKYWEALCEKILEEMREEIHRHMD-ITEDLTNKGARSGNSKPSFLAANBEIL 418
DB 2963 EVSEVI-----AE--KISEEKIEEPKPEKETEVSEKATVLDKQVL-----BEKEL 3008
QY 419 ESTRAKGDDID-----NVKSPETEKKDNENTENDSKDAEKAREE 457
DB 3009 EASAQKQGDQDVEKRSQKPEVSEVVAEKVSEGKTEEPKPEVKETEAKSE--KATTLDMQV 3067
QY 458 FEDQSLSEKSDDKTPDDDEQ--GKSEVGDPKSEKSNGLSESPGACKGASGSTRITRL 515
DB 3068 LEERELEA--SAQKQGDQDVEKKSQKPEVSEVVAEKISEEKIEEP----- 3110
QY 516 RNPDSKLSQLKSQQAHAHAANKLFKEGKEV-LVNVNSQG--EISRLSTKKEV----- 565
DB 3111 KKPEKETEVKSEK-----ATVLDKQVLEKELEASAKQGDQDVEKKSQKPEVSEVVAEK 3166
QY 566 IMGNINNYPKL-----GQEGKYR-----GQEGKYR-----VYH 587
DB 3167 VSEKIEEPKPEVKETEVEKSEKATTLDKQVLEKELEASAKQGDQDGRSDDIITLK 3226
QY 588 NOYSTNSFALNK-----HOHRE-----DHDK-----RRHLAKHFCILTAPGEKFN 627
DB 3227 ERUTELSKALGSSVDEILRESREIVNLEDDKVVAKHLFKLRDHIVHTYDGKKGGEKNK-- 3284
QY 628 GSVHGSVKVLTISTRLTITOLENNIPSSFLH-----PNWASH--RANWIKAVQCMCKPRE 680
DB 3285 -----EKELFESFIELLECEASPEAAEKVKLNLAIEK-----TN 3318
QY 681 FALALALECA-----VKPVVMLPIWRFLGHTLRHMTSISIEEKEKVKKK---EKKQ 731
DB 3319 VILTKATIQLIDDSNMFTKPSLLIP-----KLENLER--VAVKIQSETYVDKSS 3365
QY 732 EEBETHQOATWVKT-----FPVKHGVKOKGGEYRVTYGGGWSWISKTHYRVFVPLPG 786
DB 3366 EKMSISLOQSLMDFVILDDFLDDETEVLKPKIENIKITLLSDYDIIEKKDGPLLTAVING 3425
QY 787 NTWV---NYRKSLEGTKNNDENDESDKRCRSRSPKKIKIE--PDSEKDEVKGSDA--- 838
DB 3426 KINWVSQHILTIIEEVK-QLTENHDOKEKDVSNAAEDNFADEKREESQKEIKDSEAKHK 3484
QY 839 -AKGAQONEMDISKITEKKDQDKVLELSDSD----- 869
DB 3485 KSKVSEKKSIEEKLDEKKEKQTESAIDEKSKQAEVSEIYSEKITDEKAQESOKKEYKGS 3544
QY 870 -----KPKCEPEMVEDDMKTESHVNCQESSQVDVYN-VSEGE---HLRTSYK 913
DB 3545 EAKPKKAKVLEKKSIEEKLDEKKEKQTESAID-EKSQAEVSEIYSEKITDEKAQESQK 3603
QY 914 KTKSSKLDGLLERRIKQFTLEEKQRLKLEKIGKIGKIGKIGKIGKIGKIGKIGKIGKIGK 973
DB 3604 KEVKDS-----EAKPKKAKVLEKKSIEEAKLE-----DKKETQTDSDAIDEKSKQAEVS 3651
QY 974 EGCQSDSMRQEQSPNANDOPEDLIQCGSQSDSVLRMSDPSTHTNKLIPKDRVLDVSI 1033
DB 3652 E-TVSEKITDEKAQESQKEVKD-----SEAKPK---KAKVLEKKS 3689
QY 1034 RSPETKCPKONSTENDIEEK-----VSDLAS-----RGOEPTKSTKGNDFFIDQSKLA 1082
DB 3690 EEEKLEKDEKQTESAIDEKSKQAEVSEIYSEKITDEKAQESOKKEYKQSEAKPKKAKVL 3749
QY 1083 SADDIGTLICKNKKPLIQEESDTIV--SSSKSALHSSVPKSTNDROATPLSRAMDFEGKL 1140
DB 3750 EKKSIEEKLDEK-----ETQTDSDAIDEKSKQAEVSEIYSEKITD-----EKAQESQKEE 3800
QY 1141 GDSSENSTLENSDVTVSQDSSEEDMIVQNSNESISEQFRTREQDQVEVLEPKLCELVSG 1200
DB 3801 VKDSEAKPKKAKVLEKKSIEEKLDEKKEKQTESAIDE-----KSQAEVSE-IVSEKITD 3855
QY 1201 ESTGNCEDRLPVKGTETANGKKPSQOKKLEERPV--NKCSDOIKLKNT---TDKKNENRE 1255
DB 3856 EKAQESQKK-EVKGSEA---KPKKAKVLEKKSIEEKLDEK-KEKQTESAIDEKSKQAEV 3910
DB | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

Db 4390 -----EAKPKAKVLEKKSIEBEKLEKKEKQTESAIDEKSKAEVSEIVSEKI 4438
QY 1879 TSQOFTSFQPRATATVIRPNTSGSGGTTNSOVITGPIQIRPGMTVIRTPLOQSTGLKAI 1938
Db 4439 TDEK-----AQSKEEVKDESEAKPKAKV---LEKKSIEEAK 4473
QY 1939 IRTPVMPGAPQVMTQIIRGPVSTAVSAPNTVSSTPGOKSLTSATSTNIOSSASOP 1998
Db 4474 LEDKRETQDSDAIDKSO---KAEVSEIVS-EKITDEKAQES-----QKEVKDSEAKP 4523
QY 1999 PRPQO-----GOVKLTMAQLTQLTQGGHGNGLTVIQQGGTQGLQIQLIPQGVTVLP 2052
Db 4524 KAKVLEKKSIEBEKLEKKEKQTESAIDEKSKAEVSEIVSEKIT--DE 4571
QY 2053 PQQLMQAMPNGTVQRELFPLATTATTSTTTTSTTAAGTGEQROSKSLSPOMVHQ 2112
Db 4572 KAQSKEEVKDE-----SEAKPKAKVLEKKSIEBEKLEKKEKQTESAI 4616
QY 2113 DKTLPAPSSVYGPAPAKAPQTAQPS-----ARPOQOTQPSQPAQPEVQTOPEVQ 2163
Db 4617 DEKSKAEVSEIVSEKITDEKAQESQMEVVDSEAKPKAKVLEKKSIEBEKLEKKEKQTES 4676
QY 2164 T-----TVSSHVPSEAPHAOSSKPOVAQSQPOSNVQSQPVRSQVRSQRI 2212
Db 4677 TDSAIDEKSKAEVSEIVSEKITDEKAQESQKEEVKDESEAK-----PKKAKV 4723
QY 2213 ---RSTPSQSPGQSQSQVQTT---TSQPIQIPTSLSIQPSGQPSQO-POVQSSSTOTL 2265
Db 4724 LEKKSIEBEKLEKKEKQTESAIDEKSKAEVSEIVSEKITDEKAQESQKEEVKDESEAKP 4783
QY 2266 SSGQTLNVQSSSPRQIQOQPOQVIAVQLOQOVQVLSQISQVVAQTAQOQS--- 2322
Db 4784 KAKVLEKKSII---BEKLEKKEKQTESAIDEKSKAEVSEIVSEKITDEKAQESQK 4839
QY 2323 -----GVP-----QIKQLPQIQOQSSAV---QTHQIQNVTVVQAASVQEO 2361
Db 4840 EYKDESEAKPKAKVLEKKSIEBEKLEKKEKQTESAIDEKSKAEVSEIVSEKITDEKAQ 4899
QY 2362 LORVOQLRQOQKKQOQOQ-----ETKREHTLOASQSEIQQVVMKHNATIEHLQ 2414
Db 4900 ESQKEEVKDESEAKPKAKVLEKKSIEBEKLEKKEKQTESAIDEK---FQKAEVSEIVSE 4956
QY 2415 KKSMTPAEENORMIVCNQVMK-----YILDK--IDKEKQAAKRRKREESVEQKRK-- 2466
Db 4957 K--ITDEKAESRKEEVKDESEAKPKAKVLEKKSIEBEKLEKKEKQTESAIDEKSKAE 5014
QY 2467 -----QNAKLSALLFKHKEQLRAEILKRLALDKDLQIEVQOELKDLTKKEKDLQ 2520
Db 5015 VSETVSEKITDEKAQESQKEEVKDESEAKPKAKVLEKKSIEBEKLEKKEKQ-----LDEKKEKQ--- 5066
QY 2521 LAQATAVAAPCPVTPVLPAPPAPPPPPPPGVQHTGLLSTPTLTPVASQKRKREBEKDS 2580
Db 5067 --TETKVATDKSQT-----VEVSEIVLEKISSEKAEESQKVELKDS 5106
QY 2581 SKSKKKKMMISTSKETKDKTKLYCICKTPYDESIFYGCDRCQNWYHGRVCVGLIQSEAE 2640
Db 5107 EAKSKKAKVLE-----KKST-----LKEKLENDKKQKEDGATN-----RSQKAE 5146
QY 2641 LIDEVVCPOQSTEDAMTVLTPLETKDYBGLKRLVLSQAHKMWPFLPEVDP----- 2692
Db 5147 AAD--VVPKISSEKVAEIKTP-----EPMSKAKSKRPD 5178
QY 2693 --PNADPYGYIKPEPMDLATMEERVQRYRYEKLTFEADMTKIFDNCRRYNNPSPSPFYQ 2750
Db 5179 GLPADEKSHGAKVSESVVPKNEAKTDQLSAKKPVLVDELVV-----PKRKP-VL 5228
QY 2751 CAEVLESFVQKLKGF-----KASRSNNK 2775
Db 5229 AEQTADSLQTYKMSDSEYKDKESRSARK 5260

RESULT 13
Q8MLD9

ID Q8MLD9 PRELIMINARY: PRT: 9270 AA.
AC Q8MLD9;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CG18255-PA.
GN STRN-MLCK OR CG8304 OR CG18255.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Boutler H., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Haris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svitskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

Query Match		4.1%; Score 580.5; DB 5; Length 7210;
Best Local Similarity		18.2%; Pred. NO. 6e-16;
Matches		574; Conservative 533; Mismatches 1160; Indels 885; Gaps 133;
QY	4	EEEE-----EEDGDAEETQDSEDEDEM-----EEDDDSDYPMEEDDDDDASYCT- 51
DB	2614	EEELKASAKKQGDQIEQSKQPEVSEVVAEKISGKIIEPKKPEM-DTEAKSEKATV 2672
QY	52	-----ESSFRSHSYSTSPGRRKPRV-----HRRSP----- 78
DB	2673	LDKQVLEKELEASAQKQGDQDVEKKSQKPEVSEVVAEKISEETIEPKKPEVKDTEIKS 2732
QY	79	-----ILEEKDIPPLEFPKSSSEDLAVPNE--HIMNVIAIYEVLRNFGTVLRLSPPR 127
DB	2733	EKATALDKQVLEKELEASAQKQGDQDVEKKSQKPEVSEIVA-----EKISEKT 2781
QY	128	FEDFCAALYSQEQTILMAEMHVLLKAVLREE--DTSNTFGPADLKDSVNSTLYIFIDGM 185
DB	2782	IEEPKPEVKDTE--IKSEKATALDKQVLEKELEASAQKQGDQDVEKKSQK----- 2831
QY	186	TWPEVLRVYCESDKYHHVLPYQEAEDYPYGPVENKIKVLQFLVQDQFLTNTIAREELMSE 245
DB	2832	--PEVSEVVAEKISE-----ETIEEPKPEVKETE-----KSEKATVLDK 2870
QY	246	GVIOYDDHCRVCHKGLDLCCECTCSAVYHLECVKPPLEVPEDWQCEV--CVAHKVPGV 303
DB	2871	QVLEKELEASAQKQGD-----QDVEKFKQFAEYSEVVAEKISE 2910
QY	304	TDCVATIQNKPIRHEPIGYDRSRKYWFLNRLRIIEEDTE-----NENEKKIWIYSTKV 359
DB	2911	T--IEE--PKKPEVKDTEIKSEKATA---LDKQVLEKELEASAQKQGDQDVEKKSQRP 2962
QY	360	QLAELIDCKQVWEAEKILMEEREITHRMD--ITEDLTNKAAGNSKSFLLAAANEEL 418
DB	2963	EVSEVI-----AE--KISEKIEEPKPEKETEVKSEKATVLDKQVL---BEKEL 3008
QY	419	ESIRAKKGDID-----NVKSPETEERKDNETENDSKDAENREE 457
DB	3009	EASAQKQGDQDVEKKSQKPEVSEVVAEKVSEKIEPKPEVKETEAKSE--KATTLDMQV 3067
QY	458	FEDQSLKSDDDKTPDDQPEQ--GKSEVGDFKSEKSNGLSESPGAGKASGSTRITRL 515
DB	3068	LEERELEA--SAQKQGDQDVEKKSQKPEVSEVVAEKISEEKIEP----- 3110
QY	516	RNPDSKLSQKSOQVAAAHAANKLFKEGKEV--LVVNSOG--EISRLSTKKEY----- 565
DB	3111	KKPEKETEVKSEK---ATVLDKQVLEKELEASAQKQGDQDVEKKSQKPEVSEVVAEK 3166
QY	566	IMKGNINNYFKL-----COEGKYR-----VYH 587
DB	3167	VSEKIEEPKPEVKETEYKSEKATVLDKQVLEKELEASAQKQGDQDGRKSDRIITKLK 3226
QY	588	NOYSTNSFALNK-----HQHRE-----DHDK-----RRHLAKFCCLTPAGEFKWN 627
DB	3227	ERLTLSKALGSSVDEILRESREIVNNLEDKVVAKHLFKLRDHIVHTYDGRKEENK-- 3284
QY	628	GSVHGSKVLITSLRITITOLENNIPSSFLH-----PNWASH--RANWIKAVOMCSKPKE 680
DB	3285	-----EKELFESPIELCEASPEAAEKVKIYNLKEIK-----TN 3318
QY	681	FALALAILCA-----VKPVMLPIWRBFLGTHRLHRMTSITEREKEVKKK---EKKQ 731
DB	3319	VILTKATIQLIDDSNMFTKPSLLIP-----KLLNLER--VAVKIQSETYVDKSS 3365
QY	732	EEETMQQATWVKYT-----FPVKHQVWKQGEYRVTYGGWMSWISKTHVRFVFKPLPG 786
DB	3366	EKMISLQQLMDFIVLDDFDDTEVLPKPIENIKTLLSDYDYLEKKDGPLLTAVING 3425
QY	787	NTNV-----NTRKSLGTYKNMNMDESDRKRSRSPKKIKIE--PDSEKDEWKGSDA--- 838
DB	3426	KINWVSHILITLITIEEVK--QLTENHDQKEDKVDVNAEADNFADEKRESQKEEIKDSEAKHK 3484
QY	839	-AKGADONEMDISKITEKKDQDVKELLDSDS----- 869

DB	3485	KSVSEKKSIEEKLDEKKEKQTESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKG 3544
QY	870	-----KPCKEEPMVDDMKTESHVNCQESSQVDVYN--VSEGF---HLRTSYK 913
DB	3545	EAKPKKAKVLEKKSIEEKLDEKKEKQTESAID--EKSQKAEVSEIVSEKITDEKAQESQK 3603
QY	914	KKTKSKLGLLERRIKQFTLEEKQRLKIKLGGIGKIGKGTSTNSKNLSESPVTKAK 973
DB	3604	KEVKDS-----EAKPKKAKVLEKKSIEEAKLE-----DKKETQDSDAIDEKSQKAEVS 3651
QY	974	ECCOSDSMRQEOQSPNANNQDPELIOGCCSDSSVLRRMSDPSTHTNKLYPKDRVLDVSI 1033
DB	3652	E-TVSEKITDEKAQESQKEVKD-----SEAKPK---KAKVLEKKS 3689
QY	1034	RSPETCKPKONSTENDIEEK-----VSDLAS-----RGOEPTKSTKGNDFIDSKLA 1082
DB	3690	EEBLEDKKEKQTESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKDSEAKPKKAKVL 3749
QY	1083	SADDITGLCKNKKPLIOESDTIV---SSKSALHSVPKSTNDRDATPLSRAMDFEGLK 1140
DB	3750	EKKSIEEKLDEK---ETQDSDAIDEKSQKAEVSEIVSEKITD-----EKAQESQKEE 3800
QY	1141	GDSESNSTLENSSDTVSITODSSEEDMIYONSNESISEQPTREODVEVLEPLKCELVSG 1200
DB	3801	VKDSEAKPKKAKVLEKKSIEEKLDEKKEKQTESAIDE---KSQKAEVSE--IVSEKITD 3855
QY	1201	ESTGNCEDRLVPVKTEANGKKPSQOKKLEBEPV--NKCSDOIKLKNT---TDKKNENRE 1255
DB	3856	EKAQESQK--EYKGSSEA---KPKKAKVLEKKSIEEKLDEK--KEKQTESAIDEKSQKAEV 3910
QY	1256	SE-----KKQRTSTFQINGKDNKPIYKGECLKEISESRVSVGNVPEKVNINKI 1307
DB	3911	SEIVSEKITDEKAQESQMBEVKDSEAKPK---KAKVLEKKSIE-----EERLENKKEK 3960
QY	1308	IPENDIKSTVRESAIRPFINGDVIMEDFNERNSSSETKSHLLSSSDAEGNYRDLSE--TL 1365
DB	3961	QTESAIDEKSQKAEV-----SEIVSEKITDEKAQESQKKEVKDSEAKPKKAKVLEKKS 4014
QY	1366	PSTKESDSTQTTTPSPACSPESNV-----NOVEDMEIETSEVKKVTSPITSEESNL 1418
DB	4015	EEBLEDKKEKQTESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKDSEAKPKKAKVL 4074
QY	1419	SNDFIDENGLPINKNENNGESKRVITVETVMTSTVATESKTVTKVEGDKQTVVSS 1478
DB	4075	EKKSIEEKLDEKQETQDSDAIDEKSQKAEV---SEIVSEKITDEKAQESQKEEYKDS 4130
QY	1479	ENCAKSTV---TTTTTTVTKLSTPSTGGSDIISVKEOSKTVVTTVTDTSLTTTGGTLV 1534
DB	4131	AKPKKAKVLEKKSIEEKLDEKKEKQTESAIDEKSQKAEVSEIVSEINIDE----- 4181
QY	1535	TSMTVSKYSTRD-----KVKLM-----KFSRPKTRSGTALPSYRKFTKSTKS 1580
DB	4182	-RAQESQKKEVKDSEAKPKKAKVLEKKSIEEKLDEKKEKQTESAIDE---KSQKAE 4234
QY	1581	IFVLPNDLL--KKLARKGGIGREVPYFNYNAPALDIMPYSPRPTFGITWRYRLQTVKSLA 1639
DB	4235	VSEIVSEKITDEKAQESQKKEV--DSKAKPK-----KAKVLEKKSIE 4275
QY	1640	GVSLMLRLLWASLRWDMAAKVPPGGGSTRTETSETTEITTEIKRRDQVPGYIREYECI 1699
DB	4276	EAKL-----EDKKEI--QDSDAIDEKSQKAEVS-----EIVS 4305
QY	1700	KIICPIGVPETPKETPTPQRKGLRSSALRPKRPETPKQTPGVPIETWVAEELEWEIR 1759
DB	4306	EKI-----TDEKAQESQKEVKDSEAKPKKAK-----VLEKKSIEEKL 4345
QY	1760	AFARVEKEKAQAVEQQAQRLEQQRPVIASTTSTSTSTSTISPAQKVMVAPISGSV 1819
DB	4346	---NKKQTESAIDEKSQK-----AEVSEIVSEKITDEKAQESQKKEVKG-- 4389
QY	1820	TTGTAKVLTTKVGSATVTFQONKNPHOTFATVWVKOGQSGVVOQKV--LG1IPSSTG 1878

```
Db 498 VQ--MPDGKHLH-----VLTSTSSAGQGNKMKVPKIP----- 529
Qy 2118 PAQSSSVGPAQAQAPOTAPQAPQOTQPOPAQPEVOTQEVOTQTTVSSHVPSEAOPT 2177
Db 530 --ASTSSSPATSSAQT--TNPVTPVVKQIAVKHVTKNSTQSIASSSRVALP 578
Qy 2178 HAQSKPKVAQSQPSQNVQGSQVRVQSPSQ-----TRIRPSPSPQLS 2221
Db 579 LAQIKNKLLAQOQOQST--SSSPATSSSPQKIVSKVNVNTSTSGTQLQOVFVQSGSKLV 636
Qy 2222 PQQSOVOTTTTQPIPIQHTSLQIPSGQPSQVQSSQTQLSSGOTLNOVSVSPSR 2281
Db 637 VQONAQGGKVIISTSAQAQOQGTSPVQOQOLVQSQP--IQQSPQOIS--MTQQQIVVGG--- 690
Qy 2282 POLQIQOQPQVIAVPOLOQOVVLSQTSQVVAQIQAOQSGVPOQIKLQPLQIQOQSSA 2341
Db 691 -ORITLSP--GQITVTRNVNPSQALQVQOQIQTQOQOQHHV-----VQPOQOQFV 739
Qy 2342 VOTHQIQNVTVQAASVQEQLRVQOQDRQOQKQOQIEKREHTLQASQSEIQRQV 2401
Db 740 VQSNQI-----VQ--SSPSAQTKLVQVLVQOQ--SQQTIEEKTQITTTDSNETG--TQOV 789
Qy 2402 YMKHNAVIEHLKQKSMTPAEREE-----NORMIVCNOVMKYILD----- 2441
Db 790 LVPNSTLAQOLAQAGKLOVATVNGQQVIVKPLGNQQAQIVAH--IKHOGDGNHAHIVTSNSA 847
Qy 2442 -----KIDKEEKOAAKKRKREESVQKR----- 2464
Db 848 TAVPOANQTSVPVQOALPQSPQVQVQVQOQOIHQOQSPNFBESGVPTITQOQVLTQAVQA 907
Qy 2465 -SKQATKLSALLFKHK-----EQLRAEILKK-----RALLDKDLQIE-----VQEE 2505
Db 908 PAQQAQSVESLLQNPQPGTVKICVTAQVLQTEHGRPRIVLQGLVGNDFTAQQLQVLQVTO 967
Qy 2506 LKRDLIKKE-----KDLMLQAQATAVAPCPVPTVP 2537
Db 968 VRQQLMKAQESKGLVGLPTKTYILAVQPENAVSQPPPLTPV 1010

RESULT 12
Q9V7G8 PRELIMINARY; PRT: 7210 AA.
ID Q9V7G8
AC Q9V7G8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CG18255 protein
GN STRN-MICK OR CG8304 OR CG18255.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celnik S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson D., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferrera C., Ferreira S., Fleischmann W.,
```

```
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodak A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphry L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Poillard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banson J., An H., Baldwin D., Banton J., Beeson J.W., Busam D.A.,
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez S., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman B., Carlson J.W., Celnik S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF003808; AAF58087.2; -.
DR FlyBase: FBgn0013988; Strn-Mick.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig; 13.
DR SMART: SM00409; Ig; 21.
DR SMART: SM00408; Igc2; 20.
DR PROSITE: PS50835; IG_LIKE; 20.
DR PROSITE: PS00290; IG_MHC; 1.
KW Immunoglobulin domain.
SQ SEQUENCE 7210 AA; 808965 MW; 6B2E7395C07140D2 CRC64;
```

```
QY 1188 EVLEPLKELVSGESTGNCEDRLPVKGTEANGKKPSQOKKLEBPYNKCSQJLKNYTD 1247
Db 985 -----GYLSAKQAHDERKLEE----- 1001
QY 1248 KKNENRESEKKGQRTSTFOINGKDNKPKIYLKGECLKEISERVSNGVPEPKVNNINKI 1307
Db 1002 ----- 1001
QY 1308 IPENDIKSLVTKESAIRPFINGDVIMEDFNERNSETKSHLLSSDAEGNYRDSLETPLPS 1367
Db 1002 ----- 1001
QY 1368 TKESDSQTQTTSPASCPESNVQVEDMEIETSEVKVKVTSSPITSSEESNLNDFIDENG 1427
Db 1002 ----- 1001
QY 1428 LPINKNENVNAGESKRKTIVETVMTSTVATESKTVIKVEKDKQTWVSSTENCAKSTVT 1487
Db 1002 -----SGVLGEK----- 1009
QY 1488 TTTTIVTKLSTPSTGGSDIISVKEQSKTVVTTVTDSLTTTGGTLVTSMTVSKEYSTRD 1547
Db 1010 ----- 1009
QY 1548 KVKLMFSPKPTRSGTALPSYKFKVTKSTKSFIVLPNDLKKLARKGGIREV--PYFN 1605
Db 1010 -----AMPPEIQFTSTKRGKGFIVLQKRLKILQMIMGCGCOQVYPGFS 1054
QY 1606 YNAKPALDIWPPSPRPTFGITWRYLQTVKSLAGVSLMLRLWASLRWDDMAAKVPPGG 1665
Db 1055 AGIKSNLLIWPYPAPRPTDLCKWQTLNARSLHVALQLKIWSSIKFNEFD---PDDT 1111
QY 1666 GSTRTETSET--BITTTEIIKRDVGPYIRFYCIRKICPI-GVPETPKETPTPQRKG 1722
Db 1112 HPDRRVVIDPDSHERRRIIRHKEMPPYQGYERYEMEIEIPLYDEPEEDESWSNRNG 1171
QY 1723 -----LRSALPKRPET-----PKQT-----GPIIETWVAEELE 1754
Db 1172 GSSEFSHRSSARKRPQRHEFTLSLKFNGYPKKSNAPRSLDNRRAIRREWVDGVTLK 1231
QY 1755 LWE-----IRAFARVERKEKAQAV-----EQAKKLEQOKPTVATSTTSPTSST 1800
Db 1232 VFEIKDYKWIRAEAKTARKLEATKKAQAKADEERRRQQOQSVARIIPV-PMHS- 1289
QY 1801 TSTISPAQKVMVAPISGSV---TTGTMVLLTKVGSAPVTFQON-----KNFHTQF 1849
Db 1290 ---LIPSENNV-PYLGSOQORRPNNGERGFLEKYNNSSVSPQAHGYASTPPPGYHQPQ 1345
QY 1850 ATWVKQGSNSGVVQVKVLGIIIPSTGTSQFTTSFQPRATVIRPNTSGSGTTSN 1909
Db 1346 PNTIROAGYNQ-----LPRKPTTSPNFOS-RP-VATIPTPOLRAAGADG- 1390
QY 1910 SQVITGPQIRPG---MTVIRTPLOQSTLTKGAIIRTPVMVOPCAPOQVMTIIRGO--- 1961
Db 1391 --VVRAMVTPGNKSVTNTSTPYQ-----ALNRQOYQLOROOQOPAVRLTNGHFMD 1443
QY 1962 -----PVSTAVSAPNTVSSTPGQKSLTSAT----- 1986
Db 1444 GTRMGGRNPSVQMHLQPNRAALQRPFGESTTEMRRTVEAALPDNDGDEQPPVPIPRYD 1503
QY 1987 STSNIQSSASQPPRQOQVKLTMALQTLQTOGHGNGOGLTVVIQGGQTTGOLQL----- 2042
Db 1504 PTSNFDQAQAQQHPOSRPVYSTPAQMIKRTOPGGVKH--NVILMKASDGTQKRWLKPQG 1561
QY 2043 IPOGVTVLPQPGQOLMQAAMPNGTVQREFLTPLATTATTASTTTTTSITTAAGTGQRQS 2102
Db 1562 FPPGTVI--STGQVVVYRQPTAVQORQLIT-----ATPGTRVVRIPNANGAPRQDDH 1613
QY 2103 KL-----SPQ-MQVHODKTLPP-----AQSSSVGPAKAQPTQAQPSARPO--- 2141
Db 1614 QVMRRVVQASGPRAEMVDDQGTTPPGQVRYVLOGNSGTPNVPKVVSSRGGPRGGLT 1673
```

```
QY 2142 ----PQTOPOSQAQ-----PEVOTQPEVQOTQTTVSSHPVSEAQ 2175
Db 1674 MQMVQOQOQHNPFOAHYDMPDDATGFAVSTTT---EQVPDEQQ 1713

RESULT 11
QY 960Y3 PRELIMINARY; PRT; 1022 AA.
ID Q960Y3
AC Q960Y3;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE LD301469
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY051776; AAK93200.1; -
DR FlyBase: FBgn0000541; E(bx).
SQ SEQUENCE 1022 AA; 111667 MW; 1174939B61962E63 CRC64;

Query Match 4.7%; Score 674.5; DB 5; Length 1022;
Best Local Similarity 25.5%; Pred. No. 5.1e-21;
Matches 286; Conservative 144; Mismatches 361; Indels 332; Gaps 45;

QY 1535 TSMVTSKEYSTRDKVKLMKFSRPKTRSGTAL---PSYRK-----FVTSTKTSFVLP 1585
Db 100 TSSPRGRVYLLNDAKL--YEQAVKTEDKSTITKPKSYRYPILSNFLTHKKRSLVLP 157
QY 1586 NDCLKKLARKGGTREVYPYFNYNNAKPALDIWPPSPRPTFGITWRYLQTVKSLAGVSLML 1645
Db 158 RFELLKLARLGGKSTNGFHHAAKN-TIWQYQCSRPLFTCHWSYRSTNATSLSLALQL 216
QY 1646 RLWASLRWDDMAKVPVPGGSTRTETSETTEITTEIIKRRDVGPGYREYCIKTIICP 1705
Db 217 RILWASLRWDDMLAK-PPSTDGKHQVTTDEIVTEILLEKLRHSGRYGKTSYLRKVVIP 275
QY 1706 IGVPETPKETPTPQKGLRSSALRPKPEPKOTGTVIIETWVAEELEWEIRAFARV 1765
Db 276 LEMPKTVREV-TSIRSGLR---KKRAESPQTEPOITEEWDEDEKLELWEIKFMGEKQ 330
QY 1766 EKEKAQAVEQ-QAKKLEQOKPTVIATSTTSPTSTTSIPAKVMVAP-IGSVTTGT 1823
Db 331 EKARLSAVTRSVASRQLE-----ASGNSGNTSTNGALGVAGRVQLAKLSEVDK--E 381
QY 1824 KVLVTTKVGSPATVTTQKNKFHOTFATWVKQGSNSGVVQVQOKVLG-----IIPSTGT 1879
Db 382 KMEQQLKL---QRAVHQQRK-----LVATGEITRSVTPVRKGQVIGSRVIVKPNPDT 430
QY 1880 SQQFTTSFQPRATVIRPNTSGSGTTSNSQVITGPIRPGMTVIRTPLOQSTLGKAI 1939
Db 431 TR----- 434
QY 1940 RTPVMVOPCAPOQVMTIIRIQGPVSTAVSAPNTVSSTPGQKSLTSATSTSNIQSSASQPP 1999
Db 435 -----QAQVTSVSTGCGANTAAA---ASPTVG-----GSTSTQS-NPSTSTPH 474
QY 2000 RPOQGVKQLTMAQLTQTOGHGNGOGLTVVIQGGQTTGOLQIPIQGVTVLPG--PQQL 2057
Db 475 KVQ-----LIIG-----PDGKSVSRGLNPGQQL 497
QY 2058 MQAAMPNGTVQREFLTPLATTATTASTTTTTSVSTAAGTGQRQSKLSQMVHODKTL 2117
```



```

Db      4  GPAAQPPAQAAPQP--QPQPAQPEVQTQP-----AVSSHPSETQPSQAQTSKP 55
QY      2185 QVAAQSPQNSVQSGSPVRVQSPQTRIRPSTPSQLSPGQSQVQTTSQPIPIQPHSL 2244
Db      56  LVATCQCPQSSVQSGSPVRVQSPPLIRICPSTPSQVTPGQPPQVQTASQPIPIPTSL 115
QY      2245 QIPSGQSPQSPQVQSQSTLSSGQTLNVSYSPPRQLOIQPQVIAVPOLOQQVQ 2304
Db      116 QAFSGQSPQSPQVQSQSTLSSGQTLNQVTVLSPSCP-----QPQPVIAVQQL-QQVQ 169
QY      2305 VLSQIQSVVQAQIAQSQGVPOQIKQLPIQIQOQSSAVQTHQIQNVTVVQAASVQQLOR 2364
Db      170 VLSQIQSVVQAQIAQSQGVPOQIKQLPIQIQOQSSAAQT---QSVTVVQAASVQQLOR 226
QY      2365 VQQLRQOQKQKQOIEIKREHTLOASNOSEIIQOVVMKHNVAIEHLKQKSMTPAERE 2424
Db      227 VQQLRQOQKQK--QOIEIETREHTLOASNOSEIIQOVVMKHNVAIEHLKQKSMTPAERE 295
QY      2425 ENQRMIVCNQVMKYIILDKIDKEEQAAKKRREESVEQKRSKQNAATKLSALLFKHKEQLR 2484
Db      286 ENQRMIVCNQVMKYIILDKIDKEEQAAKKRREESVEQKRSKQNAATKLSALLFKHKEQLK 345
QY      2485 AEILKRALDKDLQIEVQBELKRLKIKKEKDLMLQAQATAVAAPCPVTPVLPAP-PA 2543
Db      346 AEILKRALDKDLQIEVQBELKRLKIKKEKDLMLQAQATAVAAPCPVTPVLPAPAPA 405
QY      2544 PPSPPP--PPPGVQHTGLLSTPT--LPVASQKREKREEDSSSKKKMISTTSKTKK 2599
Db      406 APPAPRPPSPSTHSLPPAGHPAPLPVTSQKKRREKDK--SKSKKKMISTTSKREAK 463
QY      2600 DTKLYCICKTPYDESKFYIGCDRCQNWYHGRVCGILQSEAEILDEYVVCQCSQSTEDAMTV 2659
Db      464 DTRLYCICKTPYDESKFYIGCDRCQNWYHGRVCGILQSEADLIDYVVCQCSQSTEDAMTV 523
QY      2660 LTPLETKDEGLKRLVRLSQAQKMAWPFLEPVPDNDAPDYGVYIKPMDLATMEERVQR 2719
Db      524 LTPLETKDEGLKRLVRLSQAQKMAWPFLEPVPDNDAPDYGVYIKPMDLATMEERIQR 583
QY      2720 YYEKLTFEVDATKIFDNCRYNPNPSDPFFVQCAEVLESFVQKLGFKASRSNNKQST 2779
Db      584 YYEKLTFEVDATKIFDNCRYNPNRDTFFVQCAEVLESFVQKLGFKASRSNNKQST 643
QY      2780 A 2780
Db      644 A 644

RESULT 8
Q9H5E0
ID      Q9H5E0      PRELIMINARY;      PRT;      412 AA.
AC      Q9H5E0;
DT      01-MAR-2001 (TrEMBLrel. 16, Created)
DT      01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Hypothetical protein FLJ23531.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Lung;
RA      Nakajima T., Nozuchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA      Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA      Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA      Oiyasashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA      Isogai T., Sugano S.;
RL      "NEDO human cDNA sequencing project.";
RL      Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AK027184; BAB15686.1;
KW      Hypothetical protein.
SQ      SEQUENCE 412 AA; 46539 MW; D72A6D830BB12B94 CRC64;

```

```

Query Match      9.5%; Score 1368.5; DB 4; Length 412;
Best Local Similarity 57.0%; Pred. No. 3e-51;
Matches 302; Conservative 18; Mismatches 41; Indels 169; Gaps 7;

QY      2107 QMOVHOD-----KTLPPAQSSTVGVPAKAQPO-----TAQPSARPPQPTQPSQ 2149
Db      31  QOVNKGKRVNCHPCRYIKTKPCHQLSHQVWVGQKPSHRLLSQLSPSPKPSPOLSLK 90
QY      2150 AOEVEQTQP--EVQTTQTVSSHVPSEAQTHAQSSKPKQVAAQSQNSVQSGSPVRVQSPS 2208
Db      91  FRLSLFRKQLFPHMSLLKHNP---PTHSHFS-PKLOSQSPQNSVQSGSPVRVQSPS 145
QY      2209 QTRIRPSTPSQLSPGQSQVQTTSQPIPIQPHTSLSQIPSGQSPQSPQVQSQTQLSSG 2268
Db      146  QTRIRPSTPSQLSPGQSQVQTTSQPIPIQPHTSLSQIPSGQSPQSPQVQSQTQLSSG 193
QY      2269 QTLNQSVSSPSRPLQIQPQVIAVPOLOQVLSQIQSVVQAQIAQSQSGVPOQI 2328
Db      194  -----
QY      2329 KLQLPQIQOQSSAVQTHQIQNVTVVQAASVQEQLRVQQLRDOQKKQOQIEIKREHTL 2388
Db      194  -----
QY      2389 QASNQSEIIQOVVMKHNVAIEHLKQKSMTPAEREENQRMIVCNQVMKYIILDKIDKEEK 2448
Db      194  -----QVVMKHNVAIEHLKQKSMTPAEREENQRMIVCNQVMKYIILDKIDKEEK 242
QY      2449 QAAKKRREESVEQKRSKQNAATKLSALLFKHKEQLRAEILKRALDKDLQIEVQBELKR 2508
Db      243  QAAKKRREESVEQKRSKQNAATKLSALLFKHKEQLRAEILKRALDKDLQIEVQBELKR 302
QY      2509 DLKIKKEKDLMLQAQATAVAAPCPVPT---PVLPAAPPAPPPPPPPGVQHTGLLSTPTL 2565
Db      303  DLKIKKEKDLMLQAQATAVAAPCPVPTPAPPAPAPPPPPPPPPPVQHTGLLSTPTL 362
QY      2566 PVASQKKRREEEKSSSKKKMISTTSKTKKDKLYCICKTPYDESK 2615
Db      363  PAASQKKRREEEKSSSKKKMISTTSKTKKDKTKLYCICKTPYDESK 412

RESULT 9
O45409
ID      O45409      PRELIMINARY;      PRT;      1711 AA.
AC      O45409;
DT      01-JUN-1998 (TrEMBLrel. 06, Created)
DT      01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      F26H11.2a protein.
GN      F26H11.2 OR F26H11.2A.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      none;
RA      Barlow K.;
RL      Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      MEDLINE-99069613; PubMed-9851916;
RA      none;
RT      "Genome sequence of the nematode C.elegans: A platform for
RT      investigating biology.";
RT      Science 282:2012-2018(1998).
DR      EMBL; Z81515; CAB04197.1;
DR      WormPep; F26H11.2a; CEI5908.
DR      InterPro; IPR000637; AT_hook.
DR      InterPro; IPR004022; DDT_dom.
DR      InterPro; IPR001965; Znf_PHD.
DR      Pfam; PF02791; DDT; 1.
DR      Pfam; PF00628; PHD; 1.
DR      SMART; SM00384; AT_hook; 1.

```



```
DE Nucleosome remodeling factor large subunit NURF301.
GN E(BX) OR NURF301 OR CG7022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21468388; PubMed=11583616;
RA Xiao H., Sandaitzopoulos R., Wang H., Hamiche A., Ranallo R., Lee K.,
RA Fu D., Wu C.;
RT "Dual functions of largest nurf subunit nurf301 in nucleosome sliding
RT and transcription factor interactions.";
RL Mol. Cell 8:531-543(2001).
DR EMBL; AF417921; AAL1664.1; -.
DR FlyBase; FBgn000541; E(bx).
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR004022; DDT_dom.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 3.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 2669 AA; 300567 MW; 6B4925AFF489D6F CRC64;

Query Match 19.8%; Score 2839.5; DB 5; Length 2669;
Best Local Similarity 27.5%; Pred. No. 2.6e-113;
Matches 869; Conservative 383; Mismatches 919; Indels 985; Gaps 98;

QY 4 EEEEDGGAETQDSEDEDEDEDDDDSDYPPEMEDDDDDASCTESSFRSHSTYSS 63
DB ESEYHYGSGFDSEEDKSDNEDDMLTTPSDSELEVANESEFSVC---SFQNGV--- 152
QY 64 TPCRRPRVHRPRSPILKE-KDIPPLEFPKSSDLMPNPHINNVTAIYEVLRNFGTVLR 122
DB ---GRPRPPSPVWVQGRVQAALDLPDSSDDLFIANTHVLRLALSIYIEVLRFRHWR 209
QY 123 LSPFRFEDCAALVSQOCTLMAEMHVLLKAYLREEDTSNTTFFPADLKDSYNTLIYFI 182
DB LSPFRFEDCAALACEEQSALLTEVHMLLKALIREDAQGTHFGPLDQDVTNISLYLI 269
QY 183 DGMTWPEVLRYCESKEY-----HHVLPQAEADYPYGPVENKIKVLQFLVDQFLTNTIA 238
DB DSIPTWPEVLRYSVESDKTFRNPFHLSHTE---YPYTGYNLEVLQFLSDQFLTSNSI 326
QY 239 REELMSEGVIQYDDHCRVCHKGLDCCETCSAVYHLECVKPPLEVEPDEMOCEVCVAH 298
DB RDVMLQEGPIHYDDHCRVCHRLGDLCCETCPAVYHLECVDPDNDVPTEDWOCGLCRSH 386
QY 299 KVPGVTDCAEIQKNKPYIRHEPIGYDRSRKRYWFLNRLIIEEDTENENKKIWIYSTK 358
DB KVSQVDCVLPQEQGVLRHDSIGVDVRHGRKYWFIAIRIFI-EDQENFT---CWYISTT 442
QY 359 VQLAELIDCLDQWYAECLKILIEEMRETHRMIDTETLTNKAQSNKSFLLAANEIL 418
DB SKLKLILLSLDAELETRLHSQITERRDETERQMKLTETLTNKHKTKRSVI----- 494
QY 419 ESIRAKKGDIDNVKSPPEETKDNKETNDSKAEKNREPEFDOSLEKSDDKTPDDDPEQ 478
DB -----ETEQEAKN-----ELLEKEVLDEDEKD----- 516
QY 479 GKSEVGDFKSEKNGELSEPGAGKGASGSTRITRLRNPDLSQLKSOQVAAAAHEAN 538
```

```

Db 914 VQAKG----- 918
QY 1127 ATPLSRAMDEGKLGCDSESNSTLENSDSTVTSIQDSSEEDMIIVONSNESISEFRTRREQD 1186
Db 919 -----GPANVLNLVSLAKIQIVRLQFS-----QL 945
QY 1187 VELEPLKCELVSGESGNCEDRLPVKGTETANGKKPSQOKKLEBRPVNKCSDOIKLKNTT 1246
Db 946 NREKAVPRC-----YKCE-----NTNSNAVSOITQ-----NTCYSPLCLQOKAR 984
QY 1247 DKKN-----NENRESEKKGORTSTFQINGKDNKPKIYLKGBCLKEISRSRVVSGNVEPKVN 1302
Db 985 AKRELLLLLKAHTAGNSKETVAAILGAVKKPSIL-----EQLTEKRE----- 1030
QY 1303 NINKIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNNSSETKSHLLSSSDAEG----- 1356
Db 1031 -----STQVAVDDSEEGKPAESEAPLDLLDWEHARAHAV 1065
QY 1357 NYRDSL-----ETLPSTK-----ESDSTOTTPPSASCPSNSVNOVEDM-----E 1396
Db 1066 PFSDSLLETCILVDQECVTNKKIQEVNASSGNCNTPDSNTQDSKIDYIESMDVCSNVE 1125
QY 1397 IETSEYKKTVPSTPITSEESNL-----SNDFID-----ENGLPINKNEN 1435
Db 1126 IESTEDSIYTGNSGNAEDVMDPGWRKRKNQSKSYIGTKDVLDTLDKDIPLNK--- 1182
QY 1436 VNGESKRKTVITEVTMTSTVATESKTVIKVEKGDQTVVYSTENCASKSTVTTTTVTVK 1495
Db 1183 ---QNRREPIT-----ARVPKREC-----VKYERETENGNERV----- 1214
QY 1496 LSTPSTGGSV-----DIISVKEQSKTVVTVTVTDSLTITGTLVTMTVSKEYSTRDKVKL 1551
Db 1215 YTSRPRGRVYLNDAAKLVEQA-----VKTEDKSTI 1246
QY 1552 MKFSRPRKTRSGTALPSYRFVTKSKTSIFVLNDDKLARKGGIREVPYFYNNAKPA 1611
Db 1247 TK--KFSYSR-----YPLISNLFTHKKRSLVLPREFELLKLARLGKSSNGFHHAAKN 1300
QY 1612 LDTPWPSRPTGTIWRYRLQTVKSLAGVSLMLRLWASLRWDDMAKVPPGGGSTRTE 1671
Db 1301 -TIWOYQCSPLPRTCWSYTSNATSLSSIALQRLWLSCLRWDDMIAT--PPSTDGKHQV 1358
QY 1672 TSETTTTTIIRRDVGPVIRFEXCIRKICPIGVPEPTPKETPPQKGLSSALRPK 1731
Db 1359 TTDEIVTLELLKHSRGYGETSYLRRKVVIPLEMPKTVREV--TSIRSLR-----KRK 1413
QY 1732 RPETPKOTGVIIETWAELELWEITRAFAERVEKEKAQAVEQQ--AKKLEQOKPTVIA 1790
Db 1414 RAESPQTEPQITEEWVDEKLELWEIKFMGEKQKARLSAVTRSVASROLE-----A 1466
QY 1791 TSTTSPTSSTTSTISPAQKVMVAP--ISGSVTTGTKMVLTKVGSPTATVTFQKNKFNHQT 1849
Db 1467 SGSGNSNTSTNGALGVAGRVOLAPKLSDEVK--EKMEQQLKL---QRAVHQRK----- 1515
QY 1850 ATWVKOGNSGVVQOKVLG---LIISSTGTS---QOTFTSFO-----PRTATVIRP 1898
Db 1516 --LVATGEITRSVTPVKGVIGSRVIVKNPDGTTIRIIQOAVTVQVSGTGANTAAAAASP 1573
QY 1899 NTSGSGTTSNS-----QVITGPQ-----IRPGMTVIRTP-----LQOSTLGA 1937
Db 1574 TVGSGTSTQSNPSTSPHKVQIIRPDGKVSVRGLNPGQOLVQMPDGKHLVLTITSSNS 1633
QY 1938 IITPVMVOPAQPOQVMTQIIRGOPVSTAVSAPNTVSTPGQKSLTSATSTSNQSSAQ 1997
Db 1634 AGOGNKKMVKIPKASTSS-----SPAISSAQTTTNPVTPVIQIAVKHVTKNSATQSIAS- 1688
QY 1998 PPRPOQGVKLTMAOLTOLTQGHGGNGLTVVLQOGOGTGO-----LQLIPOGVTV 2049
Db 1689 -----SSRVALPLAQI-----KNLLLAQOQOQOQSTSSSPATSSSPVQKRVSKVYN 1733
QY 2050 LPFGQQLQMAAMPNGTVQRFLETPLATATATSTTTTSTTAA---GTGEORQKSLSP 2106
Db 1734 TSTSGGTLOQVVFQSGS-----KLVGQNAQOGKQVLIISTSAOQOQGSTPVQOQOLVQ 1785

```

RESULT 6

Q95VB8

ID Q95VB8

AC Q95VB8

DT 01-DEC-2001

DT 01-DEC-2001

DT 01-MAR-2003

PRELIMINARY;

PRT; 2669 AA.

(Tremblrel. 19, Created)

(Tremblrel. 19, Last sequence update)

(Tremblrel. 23, Last annotation update)

```

QY 2107 OMQVHQDKTLPAPQSS--SYGPAKAQ-----POTAQPSA 2138
Db 1786 SQPIQSS-----POQISMTQVGNQPTKVKIOQIVNTSNVQOQIVVGGORIIILSPGQTIWTV 1841
QY 2139 RPQOTQOPQSPAQPEVQTOP-----EVQTOTTVSSHVPSEAPQTHAQSSKQVAAQS 2190
Db 1842 RNVPQSALQMWQOQIQOTQOQOQHVVVQOQPFVQSNQIVQSPSAQTKLVKQLVVQQ 1901
QY 2191 QPQSNVQOQSPVRVQSPSOT-----RIRSTPSQLSPQOQSQVQTTTSPQIPHPHTSLQ 2245
Db 1902 QSOQTIEEKQIITTDNETGTQOVLVPNSTLAQQLAQGLQVATVNGQOQVIVKPLGNQ 1961
QY 2246 ---IPSGOQ-----POSQOPQVOSSTQTLSSGOTLNOVSVSSPSRQLQ 2285
Db 1962 AQIVAHIKHOGDGNHIVTNSATAVPQANPQTPSPVRQQAQLPQSPQOQVVVQ-----QOQ 2016
QY 2286 IQQPOP-----QVIAPQLOQOQVQV--LSQIQSQ-----VVAQIOA 2319
Db 2017 IHQOSPNTFSGVTPITQQPVLVTOAVQAPQAOQALSVEESLLQNPQPGTVIKCVTAQVLQ 2076
QY 2320 QQSG-----VPOQIKLQLP-----IQIQQSSA 2341
Db 2077 TEHGPRIVLQGLVGNDFTAQOQLQVQVKQOLMKAQESNGKLGVLGPTKIYLAVQPEN 2136
QY 2342 VQT-----HQIQNVTVQA---ASVQEQQLQVQQLR-----DQOQKKKQOQIE 2381
Db 2137 VQSQPPPLTPVHQSAHAQOQTNIEIDATLTATTYEANSTIKDIAINNGDDQENSKCAETE 2196
QY 2382 IKREHT-----LOASNOSE-----ILOQOV--V 2402
Db 2197 NSNITTNESFAGTSSLLGSEHDEPTNLAGLIDSETDLENKQNESFVVTGQYIQKSISNA 2256
QY 2403 MKHNAVIEHLKQK-----KSMTPABREENORMIV----- 2431
Db 2257 LKQGNLSPELEELKLVCMQKQENANSTNEWETCSRGVNEALTPSRQTDDEWKIRTS 2316
QY 2432 -----CNOVMKIYLDKDEKQAAKKKREESVEQKRKQONATKLSALLPKHK 2480
Db 2317 RRPNAMTSSQFNRLK-----KNRSKNDEVAELGEOQSQ-----LERHK 2357
QY 2481 EQLRAETLKKRALLDKLOIEVOELKRDLIKKEKDMOLAOATAVAAPCPVTPVPLA 2540
Db 2358 ELKKNLTKRSLERLNLQSEIHE---DVTKVQVRHVRPLSNA----- 2397
QY 2541 PPAPPPSPPPPGVQHTGLLSTPLV-----ASQKRKEEEK----- 2578
Db 2398 ---SPDQSENERSG---EPNLDPKRTEVQNPVPRHAGRPKKLTKRKEKLYCICRTPY 2448
QY 2579 -----DSSSKSKKKMISTTSTKETHKOTKLYCICKTPYDES 2614
Db 2449 DDTKFYVGCDCSNWFHGDVCSITEASK--KLSEFICIDCKRARETOOLYCSCRPYDES 2507
QY 2615 KEYIGCDRCNWHYHRCVGILOQFAELIDEVYPCQOSTEDAMTV-LTPLTEKDYGLKR 2673
Db 2508 QFTICDCKODWHGRCVGILOQSEAFIDEVYPCQCKNDANANMKKLTNSDVEELKN 2567
QY 2674 VLRSQAHKMAWPFLEPVDNDPDDYGVYKKEPMDLATMEERVQRRYKLTETFEVADMTK 2733
Db 2568 LIQOMQLKSAWPFMEVPDPKEAPDYKYKKEPMDLKRMEIKLESNTYTKLSEFIGDMTK 2627
QY 2734 IFNCRYPNDSPPFYQCAEVLESFFVQKLUKGR 2767
Db 2628 IFNCRYPNKPSSFYKCAEALSFYVQIKNFR 2661

```

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinhart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celinker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banton J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berniker B., Carlson J.W., Celinker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Seale S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase:
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003467; AAF47362.2; -;
DR FlyBase: FBgn0000541; E(bx).
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR004022; DDT_dom.
DR InterPro: IPR006209; EGF_like.
DR InterPro: IPR01965; Znf_PHD.
DR Pfam: PF00439; bromodomain; 1.
DR Pfam: PF02791; DDT; 1.
DR Pfam: PF00628; PHD; 3.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00249; PHD; 3.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS50016; ZF_PHD_2; 2.
SQ SEQUENCE 2669 AA; 300685 MW; 261FF7B7CEEF781B CRC64;
Query Match 19.8%; Score 2842.5; DB 5; Length 2669;
Best Local Similarity 27.6%; Pred. No. 1.9e-113;
Matches 872; Conservative 381; Mismatches 920; Indels 981; Gaps 99;
QY 4 EEEEEEDGAEETQDSEDEDEDEDDDDSDYPEEMEDDDDDDDASYCTESSFRSHSTYSS 63

QY 1612 LDTWPPSPPTGTTWRYRLQTVKSLAGVSLMLRLWLASLRWDDMAKVPPGGGSTRTE 1671
D 1301 -TIWQCSPLPRTCSYTSNATSLSSIALQLRLWCLRWDDIAK-PPSTDGKHQV 1358
QY 1672 TSETETTTTEIKRRDVGPIREYECIRKICIGVPEPKETPTQKGLRSSALRPK 1731
D 1359 TTOTEIVTELELLKRISGRYGETSYLRRKVPIPLEMPKTVREV-TSISGLR----KRK 1413
QY 1732 RPETPKQTGVIIETVWABEELELWEIFRAFAERKEKAQAVEQ-AKRLEQOKPTVIA 1790
D 1414 RAESPOTPEQITEENVDEKLEWEIFKEMGEQEKARLSAVTRSVASRQLE-----A 1466
QY 1791 TSTSTSTSTSTISPAQVWAP-LSGSVTTGTQKWLTKVSGSPATVTFQKKNHQTFF 1949
D 1467 SGSNGSNTSGALGVAGRVQLAPKLSSEYK--EKMEQQLKL---QRAVHQQRK----- 1515
QY 1850 ATWVKQSGNSGVVQOKVLG---IIPSTGTS---QOTFTSFO---PRTATVTIRP 1998
D 1516 --LVATGEITRSTVPKGVIGSRRVIVNPDGTTIRLIOQAVTQVSRGTGANTAAASAP 1573
QY 1899 NTSGSGGTTNS-----QVITGPQ-----IRPGMTVIRTP-----LQOSTLGKA 1937
D 1574 TVGSGTSTQSNPSTPHKVQIIRGPDGKVSVRGLNPGQOLVQMPDGKLVLTITTTSSNS 1633
QY 1938 IIRTPVMQPGAPQVMTQIIRQPVSTAVSANTVSSPQKSLTSATSTSNISQSSAQ 1997
D 1634 AGQGNKMKVIPKASTSS-----SPAISSAQTTNPVTPVIKQIAVRHVTKNSATQSTAS- 1688
QY 1998 PPRPQOGVKLTMAQLTQLTQGHGGNGLTVVQOGQGTGQ-----LQILPOGVTV 2049
D 1689 -----SSRALPQAQI-----KNLLLAQOQOQSTSSSPATSSSPVKIVSKVNV 1733
QY 2050 LPFGQOLQMAAMPNGTVORFLEPTLATTATTTTSTVTTAA---GTGEQRQSKLSP 2106
D 1734 TSTSGTQIQVVFQSGS-----KLIVGQNAQGVKVIISTSAQOQGTSPVQOQOLVQ 1785
QY 2107 QMVGHQDKTLPPAQS-----VGPAPK--AQPTAQPSARPQPTQPSAPQPEVOTQP 2158
D 1786 SQPIQGS-----POQISMTQOQIIVGGORILSPGIVTQNNVPSQALQMWQOQIQTOQ 1841
QY 2159 -----EVOTQTTVSSHVPSEAOPHTHAQSSKPOVAQAQSPQSNVQSGSPVRVPSQT 2210
D 1842 QOQHHVPOQOQVQVQSNQIVQSSPSAOTKLKVLQVQOQSQOQTIETKTQITTTDSNET 1901
QY 2211 -----RIRPSTPSQLSPGQOQSVQTTTSQPIQHTSLQ-----IPSQGQ----- 2251
D 1902 GTQQLVLPNSTLAQQAQGLQVATVNGQVIVKPLGNQAOIIVAHKHQGDGNAHVTS 1961
QY 2252 -----PQSQPQVQSGSTQTLSSQTLNQVSVSPSPRQIQIQOQPP----- 2291
D 1962 NSATAVPOANPQTSPPVKQALPQSPQVQVVO-----QQIHOQSPTNFESGVTPITQOP 2016
QY 2292 ---QVIATVPOLOQOVV-LSQIQSO-----VVAQIQAOQSS----- 2323
D 2017 VLTOAVQAQAQQAQALSVESLQONQPPGVIVKCTAQVILQTEHPRIVLQGLVGNDFTAQ 2076
QY 2324 ---VPQIQKLQLP-----IQIQQSSAVQPT-----HQIQ 2348
D 2077 QLQVQTVQKQQLMAQESNGKLVGLPQTKIYLAVQENAVQSPPLTPVHQSAHQQT 2136
QY 2349 NVVTVOA-----ASVQEQQLRVQQLR-----DOQKKKQOQIQIEKREHT-----LQAS 2391
D 2137 NNIEIDADTLATYEANSTIKDIAINNGDDQNSKCAETENSNTTNESFAGTSSLLEGS 2196
QY 2392 NOSE-----IIOKOV--VMKNVAVIEHLKQK----- 2415
D 2197 EHDEPTNLADISDTDLNKNQESVFTVRYIQKISNALQGNLSPELEKLCVMQKQ 2256
QY 2416 -----KSMTPAERENORMIV-----CNQVMKYL 2440
D 2257 QENANSTNEWETCSRVNEALTPSRQTDDETKTISLRPNAMTTSQFNRLIK--- 2313

QY 2441 DKIDKEEKOAAKKRKRREESVEQKRSKQNTATKLSALLFKHKEQLRAEILKRALDKDLQI 2500
D 2314 -----KNKSNDEVAELGEQKQSQ-----LERHKELLKNILRRKSLIERNLQ 2357
QY 2501 EVOEELKRLDKIKKEDLMOLAQATAVAAPCPVPTVLPAPPAPPPSPPPPGVQHTGLL 2560
D 2358 EIHE---DVKTVQHRVRLPSNA-----SPDQSENEKSG-- 2389
QY 2561 STPTLPV-----ASQKRKEEK----- 2578
D 2390 -EPNLDKFRTEVQNPVRHAGRPKKLTKRKEKLYICRTPYDDTKFYVGCDCSNWFHGD 2448
QY 2579 -----DSSSKSKKMMISTTSKTKDKTKLYCICTPYDESKFYIGCDRCQNWTHGRCVGI 2634
D 2449 VSITEEASK-KLSEFICIDCKRARETQQLYCSQCRPYDESOFTICCDKQDNWFHRCVGI 2507
QY 2635 LQSEAEILIDRYCPQOQSTEDAMTV-LTPLTEKDYGLKRVLSLOAHKMAWPFLEPVD 2693
D 2508 LQSEAEFIDEYVCPQCRKNDANAANKLTSNDVLEKMLIKOMQLKSAWPFMEPVD 2567
QY 2694 NDAPDYGVYKPEMDLATMEERVORRYEKLTEFVADMTKIFDNCRYNPSDPFFVQCAE 2753
D 2568 KEAPDYKVIKPEMDLKRMBIKLESNTYTKLSEFIDMTKIFDNCRYNPKSFYKCAE 2627
QY 2754 VLESFFVQKLKGFK 2767
D 2628 ALESYFVQIKNFR 2641
RESULT 5
Q9W0T1 PRELIMINARY; PRT; 2669 AA.
AC Q9W0T1;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE CG32478 protein.
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscophila;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischman W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jaimel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K., Nixon K., Nusskern D.R., Paclet J.M.,

QY 1503 AGAAATTTTGGATCCATAAGAGCCCAAAAGGAGAGACATTGATAATGTTAAAGCCCGA 1562
Db 1 AGAAATTTTGGATCCATAAGAGCCCAAAAGGAGAGACATTGATAATGTTAAAGCCCGA 60
QY 1563 AGAAACAGAAAAGACAGAAGTGAAGTCTGAGATGATCTTAAAGTCTGAGAAAACAG 1622
Db 61 AGAAACAGAAAAGACAGAAGTGAAGTCTGAGATGATCTTAAAGTCTGAGAAAACAG 120
QY 1623 AGAAGAAATTTGAAGACCGCTCCCTTGAAGAGACAGTGAAGCAAAACACACAGATGATGA 1682
Db 121 AGAAGAAATTTGAAGACCGCTCCCTTGAAGAGACAGTGAAGCAAAACACACAGATGATGA 180
QY 1683 CCTGAGCAAGGAAATCTGAGAGCCCAACAGAAAGTTGGGGATAAAGGTAACCTCTGTGC 1742
Db 181 CCTGAGCAAGGAAATCTGAGAGCCCAACAGAAAGTTGGGGATAAAGGTAACCTCTGTGC 240
QY 1743 AGCAATCTTGGGACACACAAACAAATGCACTTCCAGAGAGACTAGTCCCTCTGAAG 1802
Db 241 AGCAATCTTGGGACACACAAACAAATGCACTTCCAGAGAGACTAGTCCCTCTGAAG 300
QY 1803 GAGGAGCCCTGTGGGTGTCTCTCAGAAACCCCGATAGCAGCAACATGCCAGAGAA 1862
Db 301 GAGGAGCCCTGTGGGTGTCTCTCAGAAACCCCGATAGCAGCAACATGCCAGAGAA 360
QY 1863 GTGGCATCTGAGCTCCCGAGGATGTGCCAGAAACCTTAACAAGACATGTGAG 1917
Db 361 GTGGCATCTGAGCTCCCGAGGATGTGCCAGAAACCTTAACAAGACATGTGAG 415

RESULT 8

US-09-833-790-199
; Sequence 199, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indirias, Carol Y.
; APPLICANT: Fan, Ligu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 199
; LENGTH: 574
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(574)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-790-199

Query Match 3.6%; Score 359; DB 9; Length 574;
Best Local Similarity 75.7%; Pred. No. 1.2e-71;
Matches 571; Conservative 0; Mismatches 1; Indels 182; Gaps 3;
QY 2818 GAGAAGAACAGAGAAAGAAACGATGCAGCAAGCGACATGGGTAAATACACATTT 2877
Db 1 GAGAAGAACAGAGAAAGAAACGATGCAGCAAGCGACATGGGTAAATACACATTT 60
QY 2878 CCAGTTAAGCATCAGTTTGGAAACAAAAGGTGAAGTACAGATGACAGATATGTT 2937
Db 61 CCAGTTAAGCATCAGTTTGGAAACAAAAGGTGAAGTACAGATGACAGATATGTT 120
QY 2938 GGTGGAGCTGATAGTAAACTCATGTTTATAGTTTGTCTTAATTCAGGCAT 2997
Db 121 GGTGGAGCTGATAGTAAACTCATGTTTATAGTTTGTCTTAATTCAGGCAT 180

QY 2998 ACTAATGTGAATTTACAGAAAGTCGTTAGAGGAACCAAAAATAATATGATGATAATG 3057
Db 181 ACTAATGTGAATTTACAGAAAGTCGTTAGAGGA 214
QY 3058 GATGAGTCAGATAAAAATAATGTTCAAGAAATGTTCAAGAAATGTTCAAGAAATGTTCA 3117
Db 215 214
QY 3118 TCTGAAAAGATGAGGTAAAGGTTTCAGATGCTGCAAAAGGACGACCAAAATGAAATG 3177
Db 215 214
QY 3178 GATATCTCAAAGATTACTGAGAAGAACCAAGATGTTGAAGGAGCTCTTAGATTCTGAC 3237
Db 215 214
QY 3238 AGTGATAAACCCCTGCAAGGAAGAACCAATGGAAGTAGAGATGACATGAAACAGAGTCA 3297
Db 241 AGTGATAAACCCCTGCAAGGAAGAACCAATGGAAGTAGAGATGACATGAAACAGAGTCA 300
QY 3298 CATGTAATTTGTCAGAGAGACTTCTCAAGTAGATGTTGTTCAATGTTAGTGGGTTTTCAT 3357
Db 301 CATGTAATTTGTCAGAGAGACTTCTCAAGTAGATGTTGTTCAATGTTAGTGGGTTTTCAT 360
QY 3358 CTAAGGACTAGTTTACAAAAAGAAAAACAAATCATCCAACTAGATGGACTTCTTGAAGG 3417
Db 361 CTAAGGACTAGTTTACAAAAAGAAAAACAAATCATCCAACTAGATGGACTTCTTGAAGG 420
QY 3418 AGAATTTAAACAGTTTACACTGGGAAGAAAAACAGCGACTCGAAAAATCAAGTTGGAGGT 3477
Db 421 AGAATTTAAACAGTTTACACTGGGAAGAAAAACAGCGACTCGAAAAATCAAGTTGGAGGT 480
QY 3478 GGNATTTAAGGTTAT - AGGAAGACTTCTACAAATTTCTCAAAATTTCTCAAAATTTCTCTCAATCACC 3536
Db 481 GGNATTTAAGGTTATAGGAAGAAAGACTTCTACAAATTTCTCAAAATTTCTCTCAATCACC 540
QY 3537 AGTAATAA - CGAAAGCAAAAGAGGTTGTCAGAG 3569
Db 541 AGTAATAAAGCAAAAGAGGTTGTCAGAG 574

RESULT 9

US-09-918-995-3226
; Sequence 3226, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: FROM VARIOUS CDNA LIBRARIES
; FILE OF INVENTION: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3226
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-3226

Query Match 3.3%; Score 327.8; DB 11; Length 477;
Best Local Similarity 99.4%; Pred. No. 1.4e-64;
Matches 329; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 8617 ACGCCACTAACAGAGAGGATTATGAGGGTTGAAGAGGTTGCTCCGTTCTTACAGGCC 8676
Db 1 ACGCCACTAACAGAGAGGATTATGAGGGTTGAAGAGGTTGCTCCGTTCTTACAGGCC 60
QY 8677 CATAGATGGCTGGCCTTTCCTTGAACCAAGTAGACCCCTAATGATGCCACGATTTAT 8736

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21036
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(516)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-21036

Query Match 4.68; Score 452.4; DB 11; Length 516;
Best Local Similarity 98.1%; Pred. No. 4.3e-93;
Matches 456; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2132 CTGGAGCTGGAAGAGGAGCATCTGGCTCAACTCGAATCATCACCAGATTGGGAATCCAG 2191
DB 52 CTGGAGCTGGAAGAGGAGCATCTGGCTCAACTCGAATCATCACCAGATTGGGAATCCAG 111
QY 2192 ATAGCAAACTAGTCAGCTGAAGAGCCAGCAGGCGCTGCGACATGAAGCAATA 2251
DB 112 ATAGCAAACTAGTCAGCTGAAGAGCCAGCAGGCGCTGCGACATGAAGCAATA 171
QY 2252 AATTATTAGAGGGGCAAGAGTACTGTAGTTAACTCTCAAGGAGAAATTTACGGT 2311
DB 172 AATTATTAGAGGGGCAAGAGTACTGTAGTTAACTCTCAAGGAGAAATTTACGGT 231
QY 2312 TGACACCAAAAGGAGTATCATGAAGGAATATCACAATTTTAAATTTGGTC 2371
DB 232 TGACACCAAAAGGAGTATCATGAAGGAATATCACAATTTTAAATTTGGTC 291
QY 2372 AAGAAGGGAAGTATCGCGTACCACAAATCAATCACTCCAGCAATTCATTGCTTTGAATA 2431
DB 292 AAGAAGGGAAGTATCGCGTACCACAAATCAATCACTCCAGCAATTCATTGCTTTGAATA 351
QY 2432 AGCACCAGCAGACAGAACCATGATGAAGAGGCAATCTTGCACATAAGTCTGTCTGA 2491
DB 352 AGCACCAGCAGACAGAACCATGATGAAGAGGCAATCTTGCACATAAGTCTGTCTGA 411
QY 2492 CTCACGAGGAGGATTCAAATGGAACGGTCTGTCCTGATGGTCCAAAGTTCTTACCATAT 2551
DB 412 CTCACGAGGAGGATTCAAATGGAACGGTCTGTCCTGATGGTCCAAAGTTCTTACCATAT 471
QY 2552 CTACTCTGAGACTGACTATCACCCCAATTAGAAGCAACATCCCTT 2596
DB 472 CTACTCTGAGACTGACTATCACCCCAATTAGAAGCAACATCCCTT 516

RESULT 6
US-09-918-995-15144
; Sequence 15144, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15144
; LENGTH: 490

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(490)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-15144

Query Match 4.3%; Score 421; DB 11; Length 490;
Best Local Similarity 95.1%; Pred. No. 6.4e-86;
Matches 444; Conservative 0; Mismatches 22; Indels 1; Gaps 1;

QY 9238 GTTATTGAAAAAAGAAAGAAAGCAAGCAAGAAAAAGATATATGGGTCAAGTGA 9297
DB 25 GTTATTGAGACATTAGCGCAGNGCGNGGAATTCAAAAAGATATATGGGTCAAGTGA 84
QY 9298 ACTCCATGGAATGCCAGCTCTCTTCACTGAGAGAGCTGGTTAGAGTCTCACAAA 9357
DB 85 ACTCCATGGAATGCCAGCTCTCTTCACTGAGAGAGCTGGTTAGAGTCTCACAAA 144
QY 9358 AACCTTTGACTGTATTTATTTATTTGTTGCAAAAAAGAGCGCTTTTATTGCTGCCTCAT 9417
DB 145 AACCTTTGACTGTATTTATTTATTTGTTGCAAAAAAGAGCGCTTTTATTGCTGCCTCAT 204
QY 9418 TTGTCAGCTAAGTATTTTCTTATAAAATCCAGCCCGGTTACATATAATCATCTGAT 9477
DB 205 TTGTCAGCTAAGTATTTTCTTATAAAATCCAGCCCGGTTACATATAATCATCTGAT 264
QY 9478 CTTATCATGATTCCTGTAGTTAAAGTACAGAGACCTCTAGATGCTTTTCTTCTAT 9537
DB 265 CTTATCATGATTCCTGTAGTTAAAGTACAGAGACCTCTAGATGCTTTTCTTCTAT 324
QY 9538 GAAAGGAGCTGCTATGTACACATGTGCACACACACAACTGGGAATCAACAATGAGTTT 9597
DB 325 GAAAGGAGCTGCTATGTACACATGTGCACACACACAACTGGGAATCAACAATGAGTTT 384
QY 9598 ATTGTTTCATGTTAGATTAAATTAAGCTTGCATAAAGGTTGGGCTAAGTGGTCTTGGGC 9657
DB 385 ATTGTTTCATGTTAGATTAAATTAAGCTTGCATAAAGGTTGGGCTAAGTGGTCTTGGGC 443
QY 9658 TACAGACTCTGCTCCCTTGATATAACAGTACAAATTTGTCATTTACT 9704
DB 444 TACAGACTCTGCTCCCTTGATATAACAGTACAAATTTGTCATTTACT 490

RESULT 7
US-09-918-995-34912
; Sequence 34912, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34912
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(416)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-34912

Query Match 4.2%; Score 411.8; DB 11; Length 416;
Best Local Similarity 99.5%; Pred. No. 7.3e-84;
Matches 413; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20812
; LENGTH: 1148
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: 1140, 1141, 1142, 1143, 1144, 1145, 1146, 1147, 1148
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20812

Query Match 5.6%; Score 553.8; DB 12; Length 1148;
Best Local Similarity 89.8%; Pred. No. 4.4e-116;
Matches 628; Conservative 0; Mismatches 17; Indels 54; Gaps 1;

QY 2079 GATAGTAGGTGATTTCCATCGGAGAGTCCACCGGGGAGCTAAGTGAATCTCCTGGAGC 2138
Db |||||
430 GACAGTAGGTGATTTCCAAATCGGAGAGTCCACCGGGGAGCTAAGTGAATCTCCTGGAGC 489
QY 2139 TGGAAAAGAGCATCTGGCTCAACTCGATCATCACCAGATTCGCGGAATCCAGATAGCAA 2198
Db |||||
490 TGGAAAAGAGCATCTGGCTCAACTCGATCATCACCAGATTCGCGGAATCCAGATAGCAA 549
QY 2199 ACTTAGTACGTCAAGAGCAGCAGCTGGCAGCGCTGCACATGAAGCAAAATAAATTT 2258
Db |||||
550 ACTTAGTACGTCAAGAGCAGCAGCTGGCAGCGCTGCACATGAAGCAAAATAAATTT 609
QY 2259 TAAGGAGGCGAAAGAGTACTGGTAGTTAACTCTCAAGGAGAAATTTACGGTTGAGCAC 2318
Db |||||
610 TAAGGAGGCGAAAGAGTACTGGTAGTTAACTCTCAAGGAGAAATTTACGGTTGAGCAC 669
QY 2319 CAAAAAGGAAGTATCATGAAAGGAATATCAACAATTTATTTAAATTTGGTCAAGAAG 2378
Db |||||
670 CGAAAAAGGAGTGGTTCATGAAGGAAGTATCAACAATTTATTTAAATTTGGTCAAGAAG 729
QY 2379 GAAGTATCGGCTCTACCAATCAATCTCCACCAATTCATTTGCTTTGAATAAGCACCA 2438
Db |||||
730 GAAGTATCGGCTCTACCAATCAATCTCCACCAATTCATTTGCTTTGAATAAGCACCA 789
QY 2439 GCACAGAGAAGCATGATAAGAGAAGGCATCTTGACATAAAGTTCTGTCTGACTCCAGC 2498
Db |||||
790 GCACAGAGAAGCATGATAAGAGAAGGCATCTTGACATAAAGTTCTGTCTGACTCCAGC 849
QY 2499 AGSAGAGTTCAATGTAAGCGGTCTGTCCATGGGTCCAAAGTTCTTACCATATCTACTCT 2558
Db |||||
850 AGSAGAGTTCAATGTAAGCGGTCTGTCCATGGGTCCAAAGTTCTTACCATATCTACTCT 909
QY 2559 GAGACTGACTATCACCACCAATTAGAAAAACAACATCCCTTCATCCTTTTTTCATCCCACTG 2618
Db |||||
910 GAGACTGACTATCACCACCAATTAG- - - - - 932
QY 2619 GGCATCACATAGGGCAATTTGGATCAAGGCAGTTCAGATGTGTAGCAAAACCCAGAGAAAT 2678
Db |||||
933 - - - - - ACTGGATCAAGGCAGTTCAGATGTGTAGCAAAACTCAGAGAAAT 975
QY 2679 TGCATTTGGCTTTTACCATTTTGGAGTGTGCAGTTAAACCAAGTTGTGATGCTACCAATATG 2738
Db |||||
976 TGCATTTGGCTTTTACCATTTTGGAGTGTGCAGTTAAACCAAGTTGTGATGCTACCAATATG 1035
QY 2739 GCGAGAAATTTTAGGACATACCAAGTTTACACCGGATGAC 2777
Db |||||
1036 GCGAGAAATCTTTAGGACATACCAAGCTTTCTCCCTTTGTC 1074

RESULT 4

US-09-960-706-151/c
; Sequence 151, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperp
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 151
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 AA236286
US-09-960-706-151

Query Match 5.2%; Score 516.4; DB 12; Length 519;
Best Local Similarity 99.8%; Pred. No. 9.3e-108;
Matches 517; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8750 AACCTATGGACCTTGCACATCGAAGAGATGACAAGACGATATTTATGAAAGCTGA 8809
Db |||||
518 AACCTATGGACCTTGCACATCGAAGAGATGACAAGACGATATTTATGAAAGCTGA 459
QY 8810 CGGAATTTGGCGAGATATGACCAAAATTTTCTAATCTGCTTACTACAAATCCAAAGTG 8869
Db |||||
458 CGGAATTTGGCGAGATATGACCAAAATTTTCTAATCTGCTTACTACAAATCCAAAGTG 399
QY 8870 ACTCCCATTTTACCAGTGTGCAGAGTTCTCGAATCATTTTGTACAGAAATTTGAAAG 8929
Db |||||
398 ACTCCCATTTTACCAGTGTGCAGAGTTCTCGAATCATTTTGTACAGAAATTTGAAAG 339
QY 8930 GCTTCAAAGCTACAGGTCTCATACAAACAACTGCAGTCTACAGCTTCTTAAAGTTCAG 8989
Db |||||
338 GCTTCAAAGCTACAGGTCTCATACAAACAACTGCAGTCTACAGCTTCTTAAAGTTCAG 279
QY 8990 CGTGTAAACCTAAACATAAAACACAGCAAGAAATCTGTTGCTCAACTATTTTAAATTAAG 9049
Db |||||
278 CGTGTAAACCTAAACATAAAACACAGCAAGAAATCTGTTGCTCAACTATTTTAAATTAAG 219
QY 9050 GAGCCAGATGTTTTTGTAGTCAGGCTATCCTGACAGACTTGACCTAAACTTCGTTTTTATT 9109
Db |||||
218 GAGCCAGATGTTTTTGTAGTCAGGCTATCCTGACAGACTTGACCTAAACTTCGTTTTTATT 159
QY 9110 GGTCAATAACAGTCCAATATATTTCTTGGCCAAATTTGTCCACGACGACAGAAAAAGCAA 9169
Db |||||
158 GGTCAATAACAGTCCAATATATTTCTTGGCCAAATTTGTCCACGACGACAGAAAAAGCAA 99
QY 9170 AGTCAACGACCACTATCTTGTCAAGATCAGATGTTTTTACTATTTGTGCGAGAAGCGAG 9229
Db |||||
98 AGTCAACGACCACTATCTTGTCAAGATCAGATGTTTTTACTATTTGTGCGAGAAGCGAG 39
QY 9230 AAAACATTTGTTTTTGAAGAAAAAGAAAAAGAAAGCA 9267
Db |||||
38 AAAACATTTGTTTTTGAAGAAAAAGAAAAAGAAAGCA 1

RESULT 5

US-09-918-995-21036
; Sequence 21036, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.

8564 ATGAGTATCTCTCCACAGTCCAGTCAACAGAGATGCCATGACAGTGTCTCACGCCAC 8623
|||||
299 ATGAGTATCTCTCCACAGTCCAGTCAACAGAGATGCCATGACAGTGTCTCACGCCAC 358
|||||
8624 TTAACAGAGAAGGATTATGAGGGTTGAAGAGGGTCTCCGTTCCCTTACAGGCCCAATAAGA 8683
|||||
359 TTAACAGAGAAGGATTATGAGGGTTGAAGAGGGTCTCCGTTCCCTTACAGGCCCAATAAGA 418
|||||
8684 TGGCCCTGGCCCTTCCCTTGAACAGTAGAGCCCTTAATGATGACACAGATTAATATGFGTTA 8743
|||||
419 TGGCCCTGGCCCTTCCCTTGAACAGTAGAGCCCTTAATGATGACACAGATTAATATGFGTTA 478
|||||
8744 TTAAGGAACCTATGGACCTTGGCACCATGGAAGAAGAGTACAAGACCATATATATGAAA 8803
|||||
479 TTAAGGAACCTATGGACCTTGGCACCATGGAAGAAGAGTACAAGACCATATATATGAAA 538
|||||
8804 AGCTGACCGGAATTTGFGGCAGATATGACCAAAATTTTGTAACTGCTCGTTACTACAAATC 8863
|||||
539 AGCTGACCGGAATTTGFGGCAGATATGACCAAAATTTTGTAACTGCTCGTTACTACAAATC 598
|||||
8864 CAAGTGACCTCCCATTTTACCAGTGTGCAGAAAGTCTCGAATCATCTTTGTACAGAAAT 8923
|||||
599 CAAGTGACCTCCCATTTTACCAGTGTGCAGAAAGTCTCGAATCATCTTTGTACAGAAAT 658
|||||
8924 TGAAGGCTTCAAGCTAGCAGGTCTCATAAACAACACTGCAGTCTACAGCTTCTTAA 8983
|||||
659 TGAAGGCTTCAAGCTAGCAGGTCTCATAAACAACACTGCAGTCTACAGCTTCTTAA 718
|||||
8984 GTTCAGCGTGTAACTCAATCAACACAGCAAGAACTCTGGTGTCTGAACATATTTTAA 9043
|||||
719 GTTCAGCGTGTAACTCAATCAACACAGCAAGAACTCTGGTGTCTGAACATATTTTAA 778
|||||
9044 ATTAAGGACCCAGATGTTTTTGTAGTCAGGCTATCCTGACAAGACTTGAACCTTAACTTCGTT 9103
|||||
779 ATTAAGGACCCAGATGTTTTTGTAGTCAGGCTATCCTGACAAGACTTGAACCTTAACTTCGTT 838
|||||
9104 TTATTTGGTCATAACAGTCCCAATATATTTCTGGCCAAATTTGTCCACGGCAAGAAAA 9163
|||||
839 TTATTTGGTCATAACAGTCCCAATATATTTCTGGCCAAATTTGTCCACGGCAAGAAAA 898
|||||
9164 AAGCAAGTCAACGACACCATATCTGTCAAGATCAGATGTTTTTACTATTGTGCGACA 9223
|||||
899 AAGCAAGTCAACGACACCATATCTGTCAAGATCAGATGTTTTTACTATTGTGCGACA 958
|||||
9224 ACGGAGAAAACCTTTGTTTTATGAAAAAAGAAAAAGAAAAAGAAAAAAGATA 9280
|||||
959 ACGGAGAAAACCTTTGTTTTATGAAAAAAGAAAAAAGAAAAAAGAAAAAAGATA 1015
|||||

RESULT 2

US-09-925-297-357
; Sequence 357, Application US/09925297
; Patent No. US20020081659A1

GENERAL INFORMATION:

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 357
; LENGTH: 1408
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1396)

; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-357

Query Match 6.8%; Score 673.8; DB 9; Length 1408;
Best Local Similarity 97.1%; Pred. No. 1.6e-143;
Matches 675; Conservative 9; Mismatches 11; Indels 0; Gaps 0;
QY 2083 GTAGGTGATTTTCACATCGGAGAGTCCACGGGGAGCTAAGTGAATCTCTGGAGCTGGA 2142
|||||
Db 276 GTAGGTGATTTTCACATCGGAGAGTCCACGGGGAGCTAAGTGAATCTCTGGAGCTGGA 335
|||||
QY 2143 AAAGGAGCATCTGGCTCAACTCGAATCATCACAGATTCGGGAATCCAGATACCAAACTT 2202
|||||
Db 336 AAAGGAGCATCTGGCTCAACTCGAATCATCACAGATTCGGGAATCCAGATACCAAACTT 395
|||||
QY 2203 AGTCAGCTGAAGAGCAGCAGGTGGCAGCGCTGCACATGAAGCAAAATTAATTTAAG 2262
|||||
Db 396 AGTCAGCTGAAGAGCAGCAGGTGGCAGCGCTGCACATGAAGCAAAATTAATTTAAG 455
|||||
QY 2263 GAGGGCAAGAGGTACTGTAGTTAACTCTCAAGGAGAAATTTACGGTTGACGACCAA 2322
|||||
Db 456 GAGGGCAAGAGGTACTGTAGTTAACTCTCAAGGAGAAATTTACGGTTGACGACCAA 515
|||||
QY 2323 AAGGAAGTGCATGATGAAGAAATATCAACAATTAATTTAAATTTGGTCAAGAGGAAG 2382
|||||
Db 516 AAGGAAGTGCATGATGAAGAAATATCAACAATTAATTTAAATTTGGTCAAGAGGAAG 575
|||||
QY 2383 TATCGGCTTACCACAAATCAATCTCCACCAATTCATTTTGTGTAATGAAGCACCAGCAC 2442
|||||
Db 576 TATCGGCTTACCACAAATCAATCTCCACCAATTCATTTTGTGTAATGAAGCACCAGCAC 635
|||||
QY 2443 AGAAGAGCAGATGAAGAGAGGCAATCTTGACATAGTTCTGTCTGACTCAGCAGGA 2502
|||||
Db 636 AGAAGAGCAGATGAAGAGAGGCAATCTTGACATAGTTCTGTCTGACTCAGCAGGA 695
|||||
QY 2503 GAGTTCAAAATGGAACGGTTCTGTCCATGGTCCAAAGTTCTTACCATATCTACTCTGAGA 2562
|||||
Db 696 GAGTTCAAAATGGAACGGTTCTGTCCATGGTCCAAAGTTCTTACCATATCTACTCTGAGA 755
|||||
QY 2563 CTGACTATACCCCAATTAGAAAACACATCCCTTCATCTCTTTTCCATCCCACTGGGCA 2622
|||||
Db 756 CTGACTATACCCCAATTAGAAAACACATCCCTTCATCTCTTTTCCATCCCACTGGGCA 815
|||||
QY 2623 TCACATAGGCAAAATTTGATCAAGGAGGTTCAGATGTAGCAAAACCCAGAGAAATTTGCA 2682
|||||
Db 816 TCACATAGGCAAAATTTGATCAAGGAGGTTCAGATGTAGCAAAACCCAGAGAAATTTGCA 875
|||||
QY 2683 TTGGCTTTAGCCATTTTGGAGTGTGAGTGTAAACCAAGTTGTGATGTACCAATATGGCGA 2742
|||||
Db 876 TTGGCTTTAGCCATTTTGGAGTGTGAGTGTAAACCAAGTTGTGATGTACCAATATGGCGA 935
|||||

RESULT 3

US-09-814-353-20812
; Sequence 20812, Application US/09814353
; Publication No. US20030165831A1

GENERAL INFORMATION:

; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124

D**b** 239 GTCAGAAATTGGTACCATGGGGCGCTGCCGTTGGCATCTTGCAAAGTGAGGCAGAGCTCATTTG 298

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.

CC	EMBL; AJ131693; CAB40713.1; -
DR	EMBL; AB019691; BAA78718.1; -
DR	EMBL; AJ010770; CA09361.1; -
DR	EMBL; AF026245; AAB86384.1; -
DR	EMBL; AF083037; AAD22767.1; -
DR	EMBL; AC004013; AAB96867.1; ALT_FRAME.
DR	EMBL; AF091711; AAD39719.1; -
DR	EMBL; AB018346; BAA34523.1; -
DR	EMBL; AC000066; AAC60380.1; ALT_FRAME.
DR	Genew. HGNC:379; AKAP9.

DR MIM; 604001; -; C:centrosome; TAS.
DR GO: 0005813; C:cytoskeleton; TAS.
DR GO: 0005856; C:cytoskeleton; TAS.

DR GO: 0004973; F: N-methyl-D-aspartate receptor-associated pr. . .; TAS.
DR GO: 0005515; F: protein binding activity; TAS.
DR GO: 0007165; P: signal transduction; TAS.

DR GO; GO:0006832; P:small molecule transport; TAS.
DR GO; GO:0007268; P:synaptic transmission; TAS.
KW Coiled coil; Alternative splicing; Polymorphism.

FT	DOMAIN	2554	2567	PKA-R11 SUBUNIT BINDING DOMAIN.
FT	DOMAIN	164	914	COILED COIL (POTENTIAL).
FT	DOMAIN	944	1022	COILED COIL (POTENTIAL).

FT	DOMAIN	1100	COILED COIL (POTENTIAL).
FT	DOMAIN	1185	COILED COIL (POTENTIAL).
FT	DOMAIN	1253	COILED COIL (POTENTIAL).
FT	DOMAIN	1280	COILED COIL (POTENTIAL).
FT	DOMAIN	1336	COILED COIL (POTENTIAL).
FT	DOMAIN	1392	COILED COIL (POTENTIAL).

FT	DOMAIN	1434	1459	COILED COIL (POTENTIAL).
FT	DOMAIN	1585	1659	COILED COIL (POTENTIAL).
FT	DOMAIN	1857	2455	COILED COIL (POTENTIAL).

FT	DOMAIN	2544	COILED COIL (POTENTIAL) .
FT	DOMAIN	2561	COILED COIL (POTENTIAL) .
FT	DOMAIN	2603	COILED COIL (POTENTIAL) .
FT	DOMAIN	3065	COILED COIL (POTENTIAL) .
FT	DOMAIN	3092	COILED COIL (POTENTIAL) .

FT	DOMAIN	3124	3470	COILED COIL (POTENTIAL).
FT	DOMAIN	3587	3689	COILED COIL (POTENTIAL).
FT	DOMAIN	3726	3730	POLY-LEV.

FT	DOMAIN	203	292	GLN-RICH.
FT	DOMAIN	321	1010	GLU-RICH.
FT	DOMAIN	1846	2772	GLU-RICH.

FT	28	
VARSPLOC	17	Missing (in isoform 2 and isoform 3).
FT		/Ftid=VSP_004102.
FT	1637	OLQEEI -> LATRRD (in isoform 4).
VARSPLOC	1642	

[illegible]

FT	VARSPLOC	2175	2183	MISSING (in isoform 3). /FTId=VSP-004105. SADTFQKVE -> Q (in isoform 6). *transcript_id=15121 004105
FT	VARSPLOC	2175	2182	MISSING (in isoform 3).

FT		2895	2907
FT	VARSPLIC		
FT			/FIND-VSP-004106->
FT			VFGFYNNCFSTLC -
FT			(in isoform 2, isoform 3 and isoform 6).
FT			%END-VSP-004107

FT	VARSPLIC	2895	2948
FT			Missing (in isoform 5).
FT			/Tid=VSP_004107.
FT			/Tid=VSP_004108.
FT	VARSPLIC	3001	3011
FT			SMTOTUACMBP ~ NLSITMCHOUHCAADDAATFEETISU

FT	1347	VARIANT
FT	1347	K -> KC
FT	1347	/FTid=VSP_004109.
FT	1347	SLG (in isoform 6).
FT	1347	STPQFPMGAK > RESLTSQNHASNFIAIEEEFELSH
VARSPLOC	1347	

FT	76	76	CONFLICT
FT	475	475	CONFLICT
FT	76	76	CONFLICT
FT	475	475	CONFLICT

554	CONFLICT	E -> G (IN REF. 3).
554	FT	
638	CONFLICT	R -> S (IN REF. 3).
663	FT	
663	CONFLICT	N -> S (IN REF. 3).
663	FT	

FT	CONFLICT	913	H -> N (IN REF. 3).	913	H -> N (IN REF. 3).
FT	CONFLICT	956	K -> N (IN REF. 3).	956	K -> N (IN REF. 3).
FT	CONFLICT	980	OKH -> PKP (IN REF. 1 AND 2).	982	OKH -> PKP (IN REF. 1 AND 2).

FT	CONFLICT	997	Q -> P (IN REF. 1 AND 2).
FT	CONFLICT	1001	Q -> P (IN REF. 1 AND 2).
FT	CONFLICT	1020	N -> D (IN REF. 3).

FT	CONFLICT	1028	1028	V -> E (IN REF. 3).
----	----------	------	------	---------------------

THIS PAGE BLANK (USPTO)

